

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 19, 2003, 05:12:15 ; Search time 2336 Seconds
(without alignments)
4360.438 Million cell updates/sec

Title: US-09-857-613A-28

Perfect score: 1830
Sequence: 1 MATVVRIPPTISCIHHTFRS.....IEGYKDLIKFAITCRKPE 350

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n model -DEV=xlh
-O=/cgn2.1/USPRO/spool/US09857613/runat_12032003_100036_6013/app.query.fasta_1.519
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blsum62 -TRANS=humana4.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=plc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09857613 @CGN 1.1.1667 @runat.12032003.100036.6013 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGCUT -DEV.TIMECUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vl:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*

29: em.vl:*
30: em.htg.hum:*
31: em.htg.inv:*
32: em.htg.other:*
33: em.htg.mus:*
34: em.htg.pln:*
35: em.htg.rod:*
36: em.htg.mam:*
37: em.htg.vrt:*
38: em.sy:*
39: em.htgo.hum:*
40: em.htgo.mus:*
41: em.htgo.other:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1159	63.3	1369	8 AF213481	AF213481 Perilla f
2	1154.5	63.1	1047	8 AY090280	AY090280 Arabidops
3	1154.5	63.1	1303	8 AY049258	AY049258 Arabidops
4	1151.5	62.9	1350	6 AX417728	AX417728 Sequence
5	1151.5	62.9	1350	8 AF104420	AF104420 Arabidops
6	1147.5	62.7	1296	8 AY087138	AY087138 Arabidops
7	981.5	53.6	118335	8 AC006193	AC006193 Arabidops
8	758	41.4	125422	2 AP003994	AP003994 Oryza sat
9	659	36.0	343550	1 AP003587	AP003587 Nostoc sp
10	647.5	35.4	134199	1 SYCSIRP	D64004 Synechocyst
11	579	31.6	108553	8 AF527809	AF527809 Sorghum b
12	416.5	22.8	298750	1 AF005375	AF005375 Thermosyn
13	392.5	21.4	334520	1 AP003588	AP003588 Nostoc sp
14	392	21.4	930	6 AX073663	AX073663 Sequence
15	392	21.4	957	6 AX073657	AX073657 Sequence
16	392	21.4	957	6 AX417730	AX417730 Sequence
17	392	21.4	974	6 AX073659	AX073659 Sequence
18	392	21.4	145701	1 D90914	D90914 Synechocyst
19	294	15.7	63734	1 SAE414559	AJ414559 Saccharot
20	287	15.7	63734	1 AF127374	AF127374 Streptomy
21	282.5	15.4	90445	1 AF040570	AF040570 Amycolato
22	270	14.8	12637	1 AF323753	AF323753 Streptomy
23	268.5	14.7	125561	8 NC936	AL513463 Neurospor
24	261	14.3	2122	6 A60301	A60301 Sequence 1
25	261	14.3	2122	6 AR144762	AR144762 Sequence
26	247	13.5	1276	8 NTU71107	NTU71107 Nicotiana t
27	244.5	13.4	5998	1 AB032524	AB032524 Streptomy
28	242	13.2	30000	6 AX250261	AX250261 Sequence
29	241	13.2	1392	8 TA060754	U60754 Trilicium ae
30	237	13.0	1410	6 E07846	E07846 DNA sequenc
31	234.5	12.8	349498	1 AP003002	AP003002 Mesorhizo
32	232.5	12.7	1698	1 AF216283	AF216283 Actinopol
33	231	12.6	1629	8 AF042332	AF042332 Oryza sat
34	228	12.5	11714	1 AE012928	AE012928 Chlorobiu
35	227	12.4	1383	8 ZM079669	U79669 Zea mays en
36	227	12.4	40065	8 SPBC169	Z99759 S.pombe chr
37	226	12.3	840	1 AF216282	AF216282 Halorhodo
38	226	12.3	1080	8 AY113031	AY113031 Arabidops
39	226	12.3	1302	8 AF375397	AF375397 Arabidops
40	226	12.3	1497	8 AF045570	AF045570 Arabidops
41	226	12.3	97554	8 AC009978	AC009978 Zea mays
42	225	12.3	6085	6 A60304	A60304 Sequence 4
43	225	12.3	6085	6 AR144763	AR144763 Sequence
44	224	12.2	1094	8 D89131	D89131 Schizosacch
45	223.5	12.2	1267	8 NTU71108	NTU71108 Nicotiana t

RESULT 1

ALIGNMENTS

SOURCE
Arabidopsis thaliana.

ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE
1 (bases 1 to 1303)
Cheuk, R., Chen, H., Kim, C.J., Koesema, E., Meyers, M.C., Banh, J.,
Bowler, L., Carninci, P., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y.,
Ishida, J., Jiang, P.X., Jones, T., Kamiya, A., Karlin-Neumann, G.,
Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M.,
Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K.,
Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C.,
Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K.,
Davis, R.W., Theologis, A. and Ecker, J.R.

TITLE
Arabidopsis cDNA clones

JOURNAL
Unpublished

AUTHORS
2 (bases 1 to 1303)
Cheuk, R., Chen, H., Kim, C.J., Koesema, E., Meyers, M.C., Banh, J.,
Bowler, L., Carninci, P., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y.,
Ishida, J., Jiang, P.X., Jones, T., Kamiya, A., Karlin-Neumann, G.,
Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M.,
Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K.,
Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C.,
Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K.,
Davis, R.W., Theologis, A. and Ecker, J.R.

COMMENT
Submitted (22-JUL-2001) Salk Institute Genomic Analysis Laboratory
(Signal), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs: Cheuk, R., Chen, H.,
Aribidopsis Full-length cDNA' : Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.

TITLE
The Salk, Stanford, PGEC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Cheuk, R., Chen, H.,
Kim, C.J., Koesema, E., Meyers, M.C., Shinn, P., Banh, J., Bowser, L.,
Dale, J.M., Goldsmith, A.D., Jiang, P.X., Jones, T., Karlin-Neumann, G.,
Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Nguyen, M.,
Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.L., Southwick, A.,
Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S.,
Davis, R.W., Theologis, A. and Ecker, J.R.

FEATURES
Cheuk, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to
this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)
contributed equally to this work as PIs.

SOURCE
Location/Qualifiers
1..1303
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
/clone="RAFL07-18-C14(R13225)"
/note="ecotype: Columbia"
1..67
68..1114
/note="gamma-tocopherol methyltransferase"
/codon_start=1
/product="Atlg64970/r13011_27"
/protein_id="AAK83600.1"
/db_xref="GI:15146234"
/translation="MKATLAAPSSLTSLPYRNSSFGSKSLILFPSSSSSVSMTTT
RGNVAAVAATSTELAKGIAEFYENRSGLEIETGDMHNGFTYDSDSGHK
EAOIRMTESLRPAGYDDEBEKTKKKYDVCGGSSRLTASFGAEICITLSPV
OAKRANGLIAAOSLAHKSQVADALDOPEDGKRLVSMESGDMKDAFVLELY
RVAAPGGRIILVTCHRNLSAEELADPEWDNIIDKICTFLPLPAMCSTDDVNLLOS
HSLDIDKADSENVAPFMPAVIRALTATWKLVSILRSKMSIKALTMPLMEICYKK
GVIRFGITTCOKPL"

5'UTR
CDS
3'UTR
BASE COUNT
ORIGIN
352 a 267 c 307 g 377 t

Alignment Scores:
Pred. No.: 5,57e-92 Length: 1303
Score: 1154.50 Matches: 221
Percent Similarity: 75.438 Conservative: 30
Best Local Similarity: 66.768 Mismatches: 65
Query Match: 63.098 Indels: 21
Gaps: 4

US-09-857-613a-28 (1-350) x AT049258 (1-1303)

QY 8 ProthrlleSercysllleHlsHsrhPheargSerGlnSerProargrhrPheala 27
DB 120 CCAACTCTTCTTCGCGTCAAGTCAATCGCTCTCTT-CGGTCCATCGCTCTCTCC 178
QY 28 Argllleargval---GlyProargSertrpAlaProileargAlaSerSer 46
DB 179 TCACTCTATGAGCAGCAACACCGTGAACCTGCGCGCTCTCTCATCTCCACT 238
QY 47 GluargGlyGluIleValleuGluGlnLysProLysLysaspLysLysleuGln 66
DB 239 GAG-----GGCTAAGA 250
QY 67 LysGlylleAlaGluPheTrpAspGluSerSerGlyLeuTrpLysnllleTrpGlyasp 86
DB 251 AAAGGAATACCGGAGGTTCTCATGAACTTCGGTTTGAGAGACAGATTGGGAGAT 310
QY 87 HisMetHisHlsGlyPheTrpAspSerSerTrpValSerLeuSerasp-----His 104
DB 311 CATATGATCATGATGCTTTATATGACCTGATCTTCTGTTCACTTCTATTTCTGCTCAC 370
QY 105 ArgAlaAlaGlnlleargmetlleGlnGluSerLeuArgPheAlaSerValser---Glu 123
DB 371 AAGGAAGCTAGATCCGTATGATTAAGACTCTCTCCGTTTCCGCTTACTATGAA 430
QY 124 GluargSerLysTrpProLysSerlleValaspValGlyCysGlylleGlySerSer 143
DB 431 GAGCAGCAGCAAAAGATAAAGAAAGATGATGATGTTGGTGGATGAGCAACCTCA 490
QY 144 ArgTrpLeuAlaLysLysPheGlyAlaTrnSerValGlylleHrLeuSerProValGln 163
DB 491 AGATATCTTCCCTTAATTTTGGAGTGAATGATGATGATGATGATGATGATGATGATG 550
QY 164 AlaGluArgAlaAsnAlaLeuAlaAlaGlnGlyLeuAlaAspLysValSerPheGln 183
DB 551 GCCAAGAGACCAATGATATCTCGCGCTGCTCAATCTCTCATTAAGCTTCTTCCAA 610
QY 184 ValAlaaspAlaLeuGlnGlnProPheSeraspGlyGlnPheaspLeuValTrpSerMet 203
DB 611 GTTCCGATGCGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 670
QY 204 GluSerGlyGluHlsMetProaspLysAlaLysPheValGlyGluLeuAlaArgValAla 223
DB 671 GAGAGTGTGAGATATGCTGACCAAGCCCACTTTGTAAGAGTGTGTAGCTGTGGCG 730
QY 224 AlaProGlyValAlaIlellellellellellellellellellellellellellellelle 243
DB 731 GCTCCAGAGAGTATGATATATATAGTACATGTTGCCAAGAAATCTATCTCGGGGAG 790
QY 244 GlnSerLeuHlsProTrpGlnGlnAspLeuLysLyslleCysAspAlaTrpTrpLeu 263
DB 791 GAAGCTTTCAGCGCTGTGGAGCAAAACATCTTGCAGAAAATCTGTAGACGTTTATCTC 850
QY 264 ProAlaTrpCysSerTrhSeraspLysValLysLeuLeuGlnSerLeuSerLeuGlnasp 283
DB 851 CCGGCTTGTGCTCCACCATGATATATATGATGATGATGATGATGATGATGATGATG 910
QY 284 lleLysSerGluLysTrpSerArgPheValAlaProPheTrpProAlaVallleargSer 303
DB 911 ATTAATGTGCGGATTTGTCAGAGACGATGATCTCTTGTGCGCTGCGGTTATAGGACT 970
QY 304 AlaPheTrpLysGlyLeuSerSerLeuLeuSerSerGlyGlnLysTrpLyslleLysGly 323
DB 971 GCATTAACATGAAAGCGCTGTGCTCTGCTTGCATGATGATGATGATGATGATGATG 1030

OY 324 AlaleuAlaMetProleuMetIleGluGlyTyrIleLysAspLeuIleLysPheAlaIle 343
 DB 1031 GCATTGACATGCCATTGATGATTTGAGGTTACAGAAAGGTCATTAGTTGGTATC 1090
 OY 344 IleThrCysArgLysPro 349
 DB 1091 ATCACTGCGCAGACCA 1108

RESULT 4
 AX417728 1350 bp DNA linear PAT 18-JUN-2002
 LOCUS DEFINITION Sequence 19 from Patent WO0231173.
 AX417728
 VERSION AX417728.1 GI:21522874
 KEYWORDS
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
 1 Ebner, M., Kunze, I. and Geiger, M.
 Improved method for the biosynthesis of vitamin E
 Patent: WO 0231173-A 19 18-Apr-2002;
 JOURNAL SONGENE GMBH & CO KGAA (DE)

FEATURES
 Source Location/Qualifiers
 1..1350
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 63..1109
 /note="coding for gamma-tocopherol methyltransferase"
 /codon_start=1
 /protein_id="CAD35425.1"
 /db_xref="GI:21522875"
 /translation="MKATLAAPSSLTSLPYRTNSSFGSKSLFRSPSSSSVSMTT
 RGNVAVAAATSTEARLKGIAEFYNETSGLEIMGDHMHGFYPPDSVOLDISGK
 EAOIMIEESLRFAGVTDEEEKIRKVVYDVGCGIGSRSRYLASKFACIGITLSPV
 OAKRANDLAASOSLSHKASFOVALDOPEDFGKDLVMSGEMHMPKAEVELY
 RYAAAGCRITIVTWHCHRNLSAGEPALDPHEONILDKICTFYLIPAMCSTDDVINLIS
 HSLDIDKADSENVAFTMPAVIRALTWKGLVSLRSMKSIKGLTLPIMIEGTRK
 GVIRFGITTCCKPL"

BASE COUNT 374 a 278 c 314 g 384 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.07e-91 Length: 1350
 Score: 1151.50 Matches: 230
 Percent Similarity: 75.43% Conservative: 31
 Best Local Similarity: 66.47% Mismatches: 65
 Query Match: 62.92% Indels: 21
 DB: 6 Gaps: 4

US-09-857-613a-28 (1-350) x AX417728 (1-1350)

OY 8 ProThrIleSerCysIleHisIleHisThrPheArgSerGlnSerProArgThrPheAla 27
 DB 115 CCACCTCTTCTTCCGCTCAAAAGTCAATCGCTTCTTT-CGGCTCCATCTCTCTCTCC 173
 OY 28 ArgIleArgVal---GlyProArgSerTrpAlaProIleArgAlaSerAlaIleSerSer 46
 DB 174 TCAGTCTCATGACGACCAACGCGTGAAGACGTGCTGCGCTGCTCATATCCACT 233
 OY 47 GluArgGlyGluIleValLeuGluGlnLysProLysLysAspLysLysLeuGln 66
 DB 234 GAG-----GGCCTAAGA 245

OY 67 LysGlyIleAlaGluPheTrpAspGluSerSerGlyLeuTrpGluAsnIleTrpGlyAsp 86
 DB 246 AAGAGATACCGAGTTTCAATGAACATTCGGGTTTGAGAGACGATTTGGGAGAT 305

OY 87 HisMetHisIleGlyPheTrpAspSerAspSerThrValSerLeuSerAsp-----His 104
 DB 115

DB 306 CATATCATCATGCGCTTTATATGACCCTGATCTTCTGTCACCTTCTGATTCGTCAC 365
 OY 105 ArgAlaIleGlnIleArgMetIleGlnGluSerLeuArgPheAlaSerValSer---Glu 123
 DB 366 AAGCAAGCTCAGATCCGATGATGATGAGACTCTCTCCGTTCCCGCGGTACTGATGA 425
 OY 124 GluArgSerLysTrpProLysIleValAspValGlyCysGlyIleGlyLysSerSer 143
 DB 426 GAGGAGAGAGAAAAAGATAAAGAAAGATGATGATGATGATGATGATGATGATGATGAT 485
 OY 144 ArgTyrLeuAlaLysLysPheGlyAlaThrSerValGlyIleThrLeuSerProValGln 163
 DB 486 AGATATCTTCCCTCAATATTTGAGCTGATGATGATGATGATGATGATGATGATGATGAT 545
 OY 164 AlaGluArgAlaAsnAlaLeuAlaIleAlaGlnGlyLeuAlaAspLysValSerPheGln 183
 DB 546 GCCAGAGAGACCAATGATCTCGGCGTCTCAATCACTCTCTATAGGCTTCTTCCCA 605
 OY 184 ValAlaAspAlaLeuGlnGlnProPheSerAspGlyGlnPheAspLeuValTrpSerMet 203
 DB 606 CTGGCGATCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 665
 OY 204 GluSerGlyGluHisMetProAspLysAlaLysPheValGlyLysLeuAlaArgValAla 223
 DB 666 GAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 725
 OY 224 AlaProGlyAlaIleIleIleIleValThrTrpCysHisArgAspLeuGlyProAspGlu 243
 DB 726 GCTCCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 785
 OY 244 GlnSerLeuHisProTrpGlnGlnAspLeuLysLysIleCysAspAlaTyrTyrLeu 263
 DB 786 GAAGCTTTCAGCGCGTGGAGCAAAACATCTTGACAAATCTGTAAGAGCTTCTATCTC 845
 OY 264 ProAlaTrpCysSerTrpSerAspTrpValLysLeuLeuIleSerLeuGlnAsp 283
 DB 846 CCGCTTGGTGGTCCACCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 905
 OY 284 IleLysSerGluAspTrpSerArgPheValAlaProPheTrpProAlaValIleArgSer 303
 DB 906 ATTAGCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 965
 OY 304 AlaPheThrTrpLysGlyLeuSerSerLeuLeuSerGlyGlnLysThrIleLysGly 323
 DB 966 GCATTACATGGAAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1025

OY 324 AlaleuAlaMetProleuMetIleGluGlyTyrIleLysAspLeuIleLysPheAlaIle 343
 DB 1026 GCATTGACATGCCATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1085

OY 344 IleThrCysArgLysPro 349
 DB 1086 ATCACTGCGCAGACCA 1103

RESULT 5
 AF104220 1350 bp mRNA linear PLN 05-JAN-1999
 LOCUS DEFINITION Arabidopsis thaliana gamma-tocopherol methyltransferase (g-TM)
 AF104220
 VERSION AF104220.1 GI:4106537
 KEYWORDS
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
 1 Shintani, D.K. and DellaPenna, D.
 Elevating Vitamin E content of plants through metabolic engineering
 Science (1998) In press
 JOURNAL Shintani, D.K. and DellaPenna, D.
 REFERENCE

TITLE Direct Submission
JOURNAL Submitted (03-NOV-1998) Department of Biochemistry, University of Nevada, Reno, NV 89557, USA
FEATURES Location/Qualifiers
source 1..1350
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
gene 1..1350
/gene="g-TMT"
63..1109
CDS
/gene="g-TMT"
/codon_start=1
/product="gamma-tocopherol methyltransferase"
/protein_id="AAD02882.1"
/db_xref="GI:4106538"
/translation="MKATLAAPSLTSLPRTNSPFSKSLIFRSPSSSSVMTT
RGNVAATASTFALRKGIAEFYNETSGIMEIYDHHHFFYPDDSVQSDGKH
EAOIRMIRESLEAFAGYTTDEEEKIRKVVDDCGGJGSSRYLASKRGACITITLSPV
QAKRANDLAAQSLSHKASFOVADLDPEDGKFDLVWSMESGEMDPKARFVELY
RYAAPGRITIVTWCIRMLSAEALQPEWONILDKICTFYLPAWCSYDDVNLLOS
HSLDIKCADMSENVAPEWPAVIRALTKGLVSLIRSGMSIKGALTMLMEGYKK
GVIRGIIITQCPV"

BASE COUNT 374 a 278 c 314 g 384 t
ORIGIN

Alignment Scores:
Pred. No.: 1.07e-91 Length: 1350
Score: 1151.50 Matches: 230
Percent Similarity: 75.43% Conservative: 31
Best Local Similarity: 66.47% Mismatches: 65
Query Match: 62.92% Indels: 21
DB: Gaps: 4

US-09-857-613A-28 (1-350) x AF104220 (1-1350)

QY 8 ProthHILeserCysIleHISLHISLTHrPheArGSeRGLnSerProArGThrPheAla 27
DB 115 CCAACTCTTCTTCCGCTCAAACTCAACGCTCTCTTT-CGGCTCCATCCTCCTCC 173
QY 28 ArgILArGVal---GLyProArGSeRTPAlaProILeArGLaSerAlaIaSerSer 46
DB 174 TCAGTCTCATGACGCAACGCTGAAACGCTGCGGCTGCTGCTACATCCACT 233
QY 47 GLuArGLyGLuILeValLeuGLnGLnLyProLySLyAspLySLySLeGLn 66
DB 234 GAG-----GGGCTAAGA 245
QY 67 LysGLyILeAlaGLuPheTyAspGLuSerSerGLyLeuTPGLuAsnILeTPGLyAsp 86
DB 246 AAGGATATAGCGGAGTTCTACATGAATCCGCTTTCGGGAGAGATTGGGAGAT 305
QY 87 HISMethIshISGLyPheTyAspSerAspSerThrValSerLeuSerAsp-----His 104
DB 306 CATATCATCATGCGCTTTATATGACCTGATCTTCTGTTCACTTCTGATTCGTGCTAC 365
QY 105 ArgAlaAlaGLnILeArGHeILeGLnGLuSerILeArGPheAlaSerValSer---Glu 123
DB 366 AAGGAGCTCAGATCCGCTGATTGAAGATCTCCGTTCCCGGCTTACTGATGA 425
QY 124 GLuArGSeRlyTrpProLySerILeValAspValGLyCysGLyILeGLyGLySerSer 143
DB 426 GAGGAGAGAGAAAAGATTAAGAAATGATGATGCGGTGGGATTGGAGAGAACTCA 485
QY 144 ArgTyLeuAlaLysLysPheGLyAlaThrSerValGLyILeThrLeuSerProValGln 163
DB 486 AGATATCTTGCCCTTAATTTGAGCTGAATGCATTGCGATTAAGCTGTTGAG 545
QY 164 AlaGLnArgAlaAsnAlaLeuAlaAlaGLnGLyLeuAlaAspLysValSerPheGln 183
DB 546 GCCAAGAGAGCCATGATCTCGCGGCTGCTCATCACTCTCTCATTAAGCTTCTCCAA 605
QY 184 ValAlaAspAlaLeuGLnGLnProPheSerAspGLyGLnPheAspLeuValTrpSerMet 203

DB 606 GTTCCGAGATCGTTGCATCAGCCATTCGAAGATGAAATTCGATCTAGTGTGCGATG 665
QY 204 GLuSerGLyGLnHISMethProAspLysAlaLysPheValGLyLeuAlaArgValAla 223
DB 666 GAGAGTGATGACATATGCTTGACAGCCCAAGTTTATAAAGGTTGATGACGCG 725
QY 224 AlaProGLyAlaILeILeILeValAlaThrTrpCysHISArGAspLeuGLyProAspGlu 243
DB 726 GCTCCAGAGAGATGAGATAATATAGTACATGCGCCATGCAATCTATCTGCGGGGAG 785
QY 244 GlnSerLeuHISProTrpGLnGLnAspLeuLeuLysLysILeCysAspAlaTyTrpLeu 263
DB 786 GAAGCTTTGGAGCGGTGGGAGCAAAACATCTTGACAAAATCTGTAGACGCTTATCTC 845
QY 264 ProAlaTrpCysSerThrAspTyValLysLeuGlnSerLeuSerLeuGlnAsp 283
DB 846 CCGGCTTGATGCTCACCGCATGATATGATCACTGCTTCAATCCATTCCTCCAGAT 905
QY 284 ILeLysSerGLuAspTrpSerArGPheValAlaProPheTrpProAlaValILeArGSeR 303
DB 906 ATTAAGTGTGCGATTGGTCAGAGAAGCTAGCTCTTTCGGCTGCGATTATACGACT 965
QY 304 AlaPheThrTrpLysGLyLeuSerSerLeuLeuSerSerGLyGLnThrILeLysGLy 323
DB 966 GCATTACATGAGAGGCGCTTGTCTCTCTCTGCTGATGATGAAGATATAAGCA 1025
QY 324 AlaLeuAlaMetProLeuMetILeGLyTyTrpLysLysAspLeuILeLysPheAlaILe 343
DB 1026 GCATTGACATGCCATTGATGATTGAAGTTTACAGAAGAGTGTCAATTGATTGATC 1085
QY 344 ILeThrCysArGLySPRO 349
DB 1086 ATCATTGCCAGAAAGCA 1103

RESULT 6
AY087138 1296 bp mRNA linear PLN 26-JUN-2002
LOCUS Arabidopsis thaliana clone 32143 mRNA, complete sequence.
DEFINITION AY087138
ACCESSION AY087138.1 GI:21405862
VERSION
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1296)
Haas,B.J., Volfovsky,N., Town,C.D., Troukhan,M., Alexandrov,N.,
Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
Full-length messenger RNA sequences greatly improve genome
annotation
Genome Biol. (2002) In press
2 (bases 1 to 1296)
Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
Feldmann,K.
Full-length cDNA from Arabidopsis thaliana
unpublished
3 (bases 1 to 1296)
Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
Feldmann,K.
Direct Submission
Submitted (11-MAR-2002) Ceres, Inc. 3007 Malibu Canyon Road,
Malibu, CA 90265, USA
This clone sequence is one of 5,000 Ceres full-length cDNAs made
available to TIGR and Genbank. The following quality assessment of
this set was done by comparison with known proteins: two percent of
the clones are estimated to be 5'-truncated; less than one percent
are 3'-truncated; approximately two percent represent alternative
splice variants, including unsplliced introns and spliced exons; one
percent may contain premature stop codons; five percent may have
frame shifts in a coding region. A sequence is considered to be
5'-truncated if it lacks the translation initiation start (ATG). A
sequence is considered to be 3'-truncated if it lacks the

C-terminal end of the encoded protein. Please note that these cDNA sequences are derived from the Ms or Laer ecotypes and therefore may contain polymorphisms when compared to sequences from Col-0. Genes carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5' sequences, selection of clones, and sequence assembly.

FEATURES

source

1..1296

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

/clone="32143"

CDS

68..1114

/codon_start=1

/product="gamma-tocopherol methyltransferase"

/protein_id="AA06496.1"

/db_xref="GI:21592747"

/translation="MKATLAAPSSLTSPYRNSSEKSSILFRSPSSSVKMTT

RGNVAVAATSTALRKGLAEFYNETSGLWEIDHMHGFPDSSVOLSDGKH

EAOIRMEESLRFAGVDDEEEKIKKKYVDCGIGSSSRYSKAFGACITLSPV

OAKRANDLAASOSLSKASFOYADALDPEDRGKDFVDSMGSEHMDKAFKVELV

RYAARGRIITVTMCHRNLSAGEALQMEONILDKTFTPLPAMCSDDDVNLLOS

HSLDDIKADMSENVAPFPAYIRALRWKGLVSLRSGMSIKGALTMPLMEIGKK

GVIRGIIITCKPL"

BASE COUNT

352 a 269 c 303 g 372 t

ORIGIN

Alignment Scores:

Pred. No.:

2.27e-91

Length:

1296

Score:

1147.50

Matches:

229

Percent Similarity:

75.43%

Conservative:

32

Best Local Similarity:

66.18%

Mismatches:

65

Query Match:

62.70%

Indels:

21

DB: 8 Gaps: 4

US-09-857-613a-28 (1-350) x AY087138 (1-1296)

QY 8 ProthrllesercyslleHisIleHsthrPheargserGlnserProargThrPheala 27
 DB 120 CCAACTCTCTTGGCGCTCAAACTCATCGCTTCTTT-CGGCTCCATCTCTCTCTCC 178
 QY 28 ArgileargVal--GlyProargSerThrPalaProileargAlaSerAlaSerSer 46
 DB 179 TCAGTCTCATGACGACACACGCGTGGAAACGCTGTGCGCTGCTGCTACATCCACT 238
 QY 47 GluargGlyGluIleValleuGlnGlnProlysLysaspLysLysleuGln 66
 DB 239 GAG-----GCGCTAAGA 250
 QY 67 LysGlyIleAlaGlnPheArgPheArgPheArgPheArgPheArgPheArgPheArgPhe 86
 DB 251 AAGAGATACCGAGTCTTCAACATGAACTTGGGTTTGGGAGAGATTTGGGGAGAT 310
 QY 87 HisMetHisHisGlyPheArgPheArgPheArgPheArgPheArgPheArgPheArgPhe 104
 DB 311 CATATGCATCATGCTTTTATGACCCGATTTCTTCTTCACTTCTGATTCGCTCAC 370
 QY 105 ArgAlaAlaGlnIleArgMetIleGlnGlnSerLeuArgPheAlaSerValSer--Glu 123
 DB 371 AAGAGATACCGAGTCTTCAACATGAACTTGGGTTTGGGAGAGATTTGGGGAGAT 430
 QY 124 GluArgSerLysThrProLysSerIleValAspValGlyCysGlyIleGlyLysSerSer 143
 DB 431 GAG 490
 QY 144 ArgTyrLeuAlaLysLysPheGlyAlaThrSerValGlyIleThrLeuSerProValGln 163
 DB 491 AGATATCTTCCCTTAATTTGGAGCGCATGATTCGATTCATCTCCACGCCCTTCAG 550
 QY 164 AlaGlnArgAlaAsnAlaLeuAlaAlaGlnGlyLeuAlaAspLysValSerPheGln 183
 DB 551 GCCAAG 610
 QY 184 ValAlaAspAlaLeuGlnGlnProPheSerAspLysGlnPheAspLeuValTrpSerMet 203

DB 611 GTTGGCGATCGCTTGGATGATGAGATGAGAGATGAGAGATGAGATGAGATGAGATG 670
 QY 204 GluSerGlyGlnHisMetProAspLysAlaLysPheValGlyGluLeuAlaArgValAla 223
 DB 671 GAG 730
 QY 224 AlaProGlyAlaIleIleIleIleValAlaThrTrpCysHisArgAspLeuGlyProAspGlu 243
 DB 731 GCTCCGAGAGAGATGATATATATAGCATGATGATGATGATGATGATGATGATGATG 790
 QY 244 GlnSerLeuHisProTrpLysGlnAspLeuLeuLysLysIleCysAspAlaTyrTyrLeu 263
 DB 791 GAAGCTTTCACCGCTGAG 850
 QY 264 ProAlaTrpCysSerThrSerAspTyrValLysLysLeuGlnSerLeuSerLeuGlnAsp 283
 DB 851 CCGGCTTGTGCTCCACGATGATGATGATGATGATGATGATGATGATGATGATGATG 910
 QY 284 IleLysSerGlyAspTrpSerArgPheValAlaProPheTrpProAlaValIleArgSer 303
 DB 911 ATTAAGTGTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 970
 QY 304 AlaPheThrTrpLysGlyLeuSerSerLeuLeuSerSerGlyGlnLysThrIleLysGly 323
 DB 971 GCATTAACATGAG 1030
 QY 324 AlaLeuAlaMetProLeuMetIleGlnGlyTyrLysLysAspLeuLysPheAlaIle 343
 DB 1031 GCATTAACATGAG 1090
 QY 344 IleThrCysArgLysPro 349
 DB 1091 ATCACTTCCAG 1108

RESULT 7

AC006193/c

LOCUS

DEFINITION

AC006193

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

REFERENCES

AUTHORS

JOURNAL

REFERENCES

AUTHORS

JOURNAL

REFERENCES

AUTHORS

JOURNAL

REFERENCES

AUTHORS

JOURNAL

REFERENCES

AUTHORS

JOURNAL

REFERENCES

AUTHORS

JOURNAL

REFERENCES

AUTHORS

JOURNAL

REFERENCES

AUTHORS

118335 bp DNA linear PLN 28-MAY-2000
 Arabidopsis thaliana chromosome I BAC F13011 genomic sequence,
 complete sequence.
 AC006193
 AC006193.3 GI:4739353
 HTG.
 KEYWORDS
 Arabidopsis thaliana.
 Arabidopsis thaliana.
 Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta,
 Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots,
 Rosidae, eurosids II, Brassicales, Brassicaceae, Arabidopsis.
 1 (bases 1 to 118335)
 Federici, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
 Altati, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P.,
 Gonzalez, A., Kremetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
 Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V.,
 Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.
 Unpublished
 2 (bases 1 to 118335)
 Federici, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
 Altati, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P.,
 Gonzalez, A., Kremetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
 Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V.,
 Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.
 Direct Submission
 Submitted (09-DEC-1998) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 3 (bases 1 to 118335)
 Federici, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
 Altati, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P.,
 Gonzalez, A., Kremetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
 Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V.,
 Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.
 Direct Submission
 Submitted (04-MAY-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA

Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
4 (bases 1 to 118335)
REFERENCE
AUTHORS
Fedeerspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altati, H., Araujo, R., Huizar, L., Kowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremensetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Lueros, S., Schwartz, J., Shinn, P., Toriumi, M., Vyotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.
TITLE
JOURNAL
Submitted (11-JUN-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
5 (bases 1 to 118335)
REFERENCE
AUTHORS
Fedeerspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altati, H., Nguyen, M., Lam, B., Southwick, A., Ecker, J., Theologis, A. and Davis, R.W.
TITLE
JOURNAL
Submitted (28-MAY-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT
FEATURES
On May 4, 1999 this sequence version replaced gi:4678192.
1..118335
Location/Qualifiers
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/map="q1147"
/clone="F13011"
1092..2311
/gene="F13011.1"
join(1092..1225,1474..2311)
/gene="F13011.1"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AAD38245.1"
/db_xref="GI:5042406"
/translation="MGQLIDPTKCTNTONERMFIESTEKLSAFORTAQTIVKLOO
RRPTKKMKMOSPEMTATRIPPDPSTGVCEDAMSVCKPMKLVNPNYSCSS
OOHOKRKAFINDINFIKYMESELCKADEIKELAEIDYKARRAEIMIKLAK
DYEEPMAREAEOMKRLFKELSEKSEKSVAKRKDLSEEROMHRLAEVLEERVVK
LMDARLFLERKLESELEARNQERERNMKPKILLERACSPARRCENPQIKRGINP
PPRVMAIRSKSEKSKLECKQVQLKILROKTPRCPLSPSP"
/complement(3352..3963)
/gene="F13011.2"
complement(3352..3963)
/gene="F13011.2"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AAD38246.1"
/db_xref="GI:5042407"
/translation="MGNCIFGGLGDEEDLLIKYKSDGCVLEFVSPVAGVSHGFS
GHALFSADVLLMKPLAHDLVPGOSTYLEPNIVSELKTFVGSCHVRNSSELSIT
PYRMSLDVYHNRVLKRSYTDVFSRNSHIRTROKERTRRRTSSKGAIMKVNLIINTE
LQILISEDRTHELLESYRAVAKGETSSITSSSENFLSVQT"
8107..9734
/gene="F13011.3"
join(8107..8140,8222..8358,8440..8489,8562..8893,
8959..9062,9144..9219,9304..9734)
/gene="F13011.3"
/note="very similar to alcohol dehydrogenase. Similar to
formaldehyde dehydrogenase. [Arabidopsis thaliana];
Location of est 94M15T7 (g112757072) and 94M15X (3.)
(g113450565)"
/codon_start=1
/protein_id="AAD38247.1"
/db_xref="GI:5042408"
/translation="METGQKVTTCRAAVWAGCEPLVMDVAVDPQRLVETRIILFT
SICHTDLAMKENEORAPYPIILHGAAGTIESGEEVEMAGDHVLPITGEGD
CRVCKRDGANLDERFVDMKRVWMTGDTGTRFTSDNKPPIYHPLNTSTFSEYVTL
ACVLKADPLEPLEKISLSCGVSTDKLMNEGVAMNADVADPASTVAIFGLANGVL
AVAEGARARGASKIIGIDINPDKFLQREAGISSEIFNPESDKAVHERVMEITEGVE

YSPFCAGSIEALREAFLESTNSGVGVYVAASPOLDPIHPMELFOGRSITASVYG
FKRQLPPEPTOCLOGLLDLEFISHQLEPFIHINEMQMLHOGKRLRLHLH"
complement(10308..11958)
/gene="F13011.4"
complement(join(10308..10729,10802..10883,11042..11125,
11208..11388,11461..11617,11723..11958))
/gene="F13011.4"
/note="Location of ests 66H377 (g112414136), 62B10T7
(g1931264), 228J15T7A (g111217083). 89% identical to
membrane related protein CP5 [Arabidopsis thaliana]"
/codon_start=1
/product="putative membrane related protein"
/protein_id="AAD38248.1"
/db_xref="GI:5042409"
/translation="MNGLSLVTFEFSHSCFSDLIYFAPLPIAAGLVGVWVRPK
VADLSDKLSNPKFLNQLPTSLTKTSSHVWVSLSEKTEGVDDDFRLMKLVE
WKDGCWLOMDMDRSTPTFSYQAMRDPKRPQPSRVFEDATPEMDKMLDE
RSKMDMLYSSSTLERCKDTGTIVGVWVKRPFECSDREYIISRLIMDAGRVYCITK
GVQYPSVPROMKPRRYDLVYSSWCIRAVSKRSGCEMTGCEVLEFPHEDMGIPMEIAK
LVROCMGCAVKKIIEPGLRAYORAKAGSLSPSMAIHNTKVSDEEPMENGSTAE
VGDRTKGNIRKILYVGAIALACTLDKGLTKAVIFGVARRFRMRKM"
13535..15717
/gene="F13011.5"
join(13535..13647,13808..14022,14535..14734,14807..15315
15402..15717)
/gene="F13011.5"
/note="Identical to Tubulin Alpha-1 Chain [Arabidopsis
thaliana] (g11135391) Location of ests 139020T7
(g112762968), 214C14T7 (g11159108)."
/codon_start=1
/product="alpha tubulin"
/protein_id="AAD38249.1"
/db_xref="GI:5042410"
/translation="MREIISIHIGAGCIVGNSCMELYCIEHICIQPDGTPMSDSTVGA
CHDAEFTFEFSISGQHVRAVFLDLEPTVIDEVRGTROLFHPQLISGEDAANN
FARGHYTVREIYDCLERKLADNCQLOGLFVNAAGVSGLSLLERLSD
FGKSKLGTFTIYSPQVSTAVVEYVSMTSLAEHTVVLNDNEATYDLCRSID
IEPTYSNLRILISOTISLTSLRFDGAINVIDEFOINIVPXPPIHMLSSYATV
SSAKAHDFPSVPEITTSVFPESNMMAKDDPRHKGIMACCIARGVYKRVNTAVAA
IKAKRTIYVDVDCPTGKGINQTPSPVGGDLAVQVACMISNTVAEVSID
HKFDLYMSRAVHVYVGEGMEGSEAREDLALEKYEVEVGGCAGDDEEGDEY
"
complement(16208..17133)
/gene="F13011.6"
complement(join(16208..16338,17040..17133))
/gene="F13011.6"
/note="Unknown protein; Location of ests 247F21T7
(g112413341)"
/codon_start=1
/protein_id="AAD38250.1"
/db_xref="GI:5042411"
/translation="MAAEPRKATAEVYKMDLEDEDEFEFEINEDMLEKEEYKVSQ
QWEDDDDDVDVDDFSRQRLKELENGTDRK"
complement(17641..19522)
/gene="F13011.7"
complement(join(17641..18251,18334..18833,19281..19522))
/gene="F13011.7"
/note="31% identical to glucan endo-1,3-beta-glucosidase
precursor ((1-3)-beta glucan endohydrolase) [Triticum
aestivum] (g111706551)"
/codon_start=1
/product="Similar to glucan endo-1,3-beta-glucosidase
precursor"
/protein_id="AAD38251.1"
/db_xref="GI:5042412"
/translation="MATRKLPEKTVVQMLKNNINKVKLEFDEDTMGALAGGLEVM
VAIPNDQLKVMTSYBRKDWKVRKNTRVVAFDGVNTTFAVANEPELISYNGSFNLT
PALANIONALNEMAGNCSYKATVPLNADVYSPASNPSPSGRRPDIIGOMTOYD
SLKANAPITITINIPPLSLTGNDPPLNANFAFDGAEPINDNGITNTVFADNFDTLVS
SLKAVNGCMPLIIEVGEWPTGDKHANASAVRYNGLPLPLGNTKNTPLAPTYIEV
YLGFLGDEAKSIACGPFERHNGIFKFDQGPFPIDLSGQSGSKFLICAOVWVPLPK
WCTFNPBEARDLKLANDIADCTFSDCTALPGLIDAGNAGNSVAFNNFPOVKNQ
DESACYFQGLATITITONISQGCNPIQIVVASASSFSSSVSLVVLVWFLSGMMF

```

gene      20970..22826
          /gene="F13011.8"
CDS       join(20970..21383,21459..21893,22659..22826)
          /gene="F13011.8"
          /note="Hypothetical Protein"
          /protein_id="A038252.1"
          /db_xref="GI:5042413"
          /translation="MASLISFSLPPKPAVRSSISAPOTOTITELEKDFKRGKIKF
          SESNNIPWELKVRNSSLKLSLSDAHVLSYKRYVWKDEGEVEVLYTDGDSRGV
          GVIYVGESEPKGSYVISCDSVSKDSDAIDALOIELSCAGVLDITYLSLVS
          MATALVKNNGKRPVTLKPGIKSGYLFKKRSGAGIKGSCYCPMPRLKAVNPES
          SEAMKSESGMGSEEGEGKGIWAVDSVTLLEKMSRYGAPAPRLKAVNPES
          KEPTIDDDKDWVRDVCGLPRIDVGQWEATLRLVHRSVNRSCRSCLRDMERN
          GNRA"
gene      complement(24442..25986)
          /gene="F13011.9"

Alignment Scores:
pred. No.:      2,5e-74      length:      118335
Score:          981.50      Matches:      231
Percent Similarity: 47.37%      Conservative: 30
Best Local Similarity: 41.92%      Mismatches: 65
Query Match:    53.63%      Indels:      227
DB:             8           Gaps:          6

US-09-857-613a-28 (1-350) x AC006193 (1-118335)

QY      8  Prothrillesercyslielhislethrpheargserglinsrproargthrpheala 27
        |||||:::      ||| |||      ::|||:::      :::
Db 99362 CCAACTCTCTTGGCGTCAAAAGTCATCGCTTCTT- CGGTCTCATCTCCCTCCGCC 99304

QY      28 Argileargval---GlyproargsertrpalaProilearglaaserlaser 46
        ::::      ::: |||      ::: |||::: |||:::
Db 99303 TCACTCTCTATGACGACAAACCGCTGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCT 99244

QY      47 GluarglgluilevalleuglulnlsProlylsaspasplylslysleucln 66
        |||
Db 99243 GAG-----GCGCTAAGA 99232

QY      67 Lysglilaleagluphetryaspgluserserglyleutrgluasnlletrpola 86
        |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 99231 AAAGCAATAGCGGAGCTTCTAATAAGAACTCGGCTTGCGGAAGAGATTGGGAGAT 99172

QY      87 Hismethishisgliphetryaspserasrthrvalserleuserasp-----His 104
        |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 99171 CATATGCATCATGCTTTATGACCTGATTCCTCTTCAACTTCTGATTCGTCAC 99112

QY      105 ArgalaalaglinileargmetileglinsrleuargpheelaserValser----- 122
        ::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 99111 AAGAGCTCAGATCCGATATGATGAGAGTCCTCCGTTTGGCGGTACTGCTTAG 99052

QY      122 ----- 122
Db 99051 CTTCCTCACTTTTTCCTCTGATATATACACTTGAGTTAAATACCGCTTTGATATATG 98992

QY      122 ----- 122
Db 98991 TTAACGAACAATAAAGAGTTAATAATACAACAATAAGACTTAAATCTTTAAACC 98932

QY      122 ----- 122
Db 98931 ACAGAAAAAACTTATTTGTCTACAGAAATGACTAGGAAATATGCTTTTGTGTAC 98872

QY      122 ----- 122
Db 98871 TCTGTACCCCATGAAGGTGTTATAGCCACAAGCATCTTAACCTTTTATTTTA 98812

QY      122 ----- 122
Db 98811 AGAACATACTTTTCTTGATTTGAATTAAGAGATGATTTGAACCTTAATGAATACATA 98752

```

```

QY      122 ----- 122
Db 98751 ATTTAAACCAAAAGTAATATAGTTAGAAATATATATTTATTTTATCTTTCG 98692

QY      122 ----- 122
Db 98691 TTTGTATTCAGTTATTTTACCTTTTAAATTAATATACATAGAACCTTGATGCC 98632

QY      123 -----Glulargserlystrprolysserileva 133
        ||||| ||| ||| ::::
Db 98631 ATGTAATGATTTGATTAAGAAAAAGATGAGAGAGAGAGAAAAAGATGAAGAAGTACT 98572

QY      133 LaspaValGlycysgllyleagllyserargtyrleualalyslsphelalath 153
        |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 98571 GGATGTTGGGTGGGTGGATTTGGAGAGACCTCAAGATATCTTCCCTTAAATTTGGAGCTGA 98512

QY      153 rserValGlylethrleuserprovalglinalargalalsnalaleualalal 173
        :::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 98511 ATGCATTGGCATTTACTCTCACGCCCTGTTCAGGCCAAGAGACCAATGATTCGCGCTGC 98452

QY      173 aglgllyleualasplysVal----- 180
        |||| ||||| |||
Db 98451 TCAATCACTCGCTCATTAAGT-ACGTTTGGATAAACATACCTTTCATTTTGGAAT 98393

QY      181 -----SerphelgluValalaaspaalaleucl 189
        |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 98392 TTCACTTAACGTATGACATTTGACCTTCTAGCTTCTCCACAGTGGCGAGTGTGA 98333

QY      189 nglnPropheseraspelglinpheasphevaltrpseretglusergllyluhisme 209
        |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 98332 TCACCCATTCGAAATGAGAAATTCGATCTAGTGTGCGATGAGAGATGCTGAGCAAT 98273

QY      209 tProaspplsyalysphevalgllylualeualargvalalalaproglialalleil 229
        |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 98272 GCGTGACAGGCCAAGTTTGTAAAAGACTGTGTACGTGTGCGGCTCCAGAGAGTAGAT 98213

QY      229 eilellevalthrtrpcysHisargaspheuglyproaspeluglnserleuHisprotr 249
        |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 98212 AATAATAGTACATGCTGCCATAGAAATCTATCTGCGGGGAGAGCTTTGCAACCGTG 98153

QY      249 pgluglnaspheuleulyslysleecysaspalatyrryleuprolatrpysserthr 269
        |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 98152 GGAGCAAAACATCTTGGAACAATCTGTAGACGTTCTATCTCCCGCTGTGCTCCAC 98093

QY      269 rSeraspfyrallylsleuuglnserleuserleucln----- 282
        |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 98092 CGATGATATGTCAACTTCTCAATCCCATCTCTCCAGGTTATTAATTCCTACCT 98033

QY      282 ----- 282
Db 98032 TTGCTGCCGAACATTTACTTAAGACTTTTACTAATACTGAATCGAATATGCT 97973

QY      283 -----Aspilleysserglusptrpserargphevalalaprophetrrpola 300
        |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 97972 GTTTTACGATATTAAGGTGCGGATTCGTGACAGAAAGTACCTTCTGCGCTCGG 97913

QY      300 alileargserlalephetrpulylsyleuserserleuusersergllylnlstr 320
        |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 97912 TTATACGGAGTCATTAACATGAGAGGCTTGTCTCTGCTGCTGCTGCTGCTGCTGCT 97853

QY      320 hrileysglivalaleualametrProleumetilegllytyrlylsysaspleuilel 340
        :::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 97852 GTATTAAAGGAGCATTTGCAATGCAATGATGATGATTAAGAGTTACAAAGAAAGTGTCATTA 97793

QY      340 ysphealalellethrCysarglyspro 349
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 97792 AGTTTGATCATCACTTCGCCAGAGCA 97764

RESULT 8
LOCUS      AP003994/c      125422 bp      DNA      linear      HTG 21-MAR-2002
DEFINITION      Oryza sativa (japonica cultivar-group) chromosome 2 clone
                  OJ111_E07, *** SEQUENCING IN PROGRESS ***, in ordered pieces.

```



```

QY 122 ----- 122
Db 16914 TAGTACTGCAAGAAATAGCAAAAGACCAAAATTGTCGTGAATTCAGCCGTACTTAC 16855
QY 122 ----- 122
Db 16854 CCTTGGAGTACGTACATGCTCAATTATTTGCTTACGTATATATTTGCTTGAATGCT 16795
QY 123 ---gluGluarGserLysTrpProLysSerIleValAspValGlyCysGlyIleGlyGly 141
Db 16794 GCACATGATGCGGAGAGAAACCAAAAGTGTAGTTGATGTGGCTGTGGCATTTGGTGGT 16735
QY 142 SerSerArgTyrIleuAlaLysLysPheGlyAlaThrSerValGlyIleThrLeuSerPro 161
Db 16734 AGCTCAAGATACCTTGGCGCAAAATAGCGAGCGCAATGCTACGGCATCAGCTTGAATCGG 16675
QY 162 ValGlnAlaGlnArgAlaAsnAlaLeuAlaAlaGlnGlyLeuAlaAspLys----- 179
Db 16674 GTGAGGCTGAAAGAGAAATGCCCTCGCGAGAGCAAGGCTATATCAGACAAAGCTGGCT 16615
QY 179 ----- 179
Db 16614 ATTTACTACTGTTTATCTCTCTTAATAAAATTCTACTGTTTATCTGATCGGGAATTAA 16555
QY 179 ----- 179
Db 16554 TCTGCGTGTAGTATGATTAATTGACATATCAGAGGTTTAAAGCTGTGATTAC 16495
QY 180 ----- 180
Db 16494 TCACCTGCTGTGACGACAACTCTGTTTAAAGAGCTCTCTTCAAGTTGGTGATGCATT 16435
QY 188 uGlnGlnProPheSerAspGlyGlnPheAspLeuValTrpSerMetGluSerGlyGlnIle 208
Db 16434 GGACAGAGCTCTTCTCTGATGGGACATTTGATCTGTGTCATGAGAGATGGGAGCA 16375
QY 208 smetProAspLys----- 212
Db 16374 CATGCGAACAAGGCGAGTAAATACTCCTCTTTTATCTTACGAAAGAAAAAGAA 16315
QY 212 ----- 212
Db 16314 ATGGGAAATGTATATCCCTCCGTTTACAGTTATTAAGACTTCTTACATTCGACGTTCA 16255
QY 212 ----- 212
Db 16254 ATAGATGTTAATGAATCTAGGACACATATATGTTTACATTTACATATATATGAAT 16195
QY 212 ----- 212
Db 16194 ATGGGTATACTAGAAAGTTTATATATGAAGGAAAGAAATATCTTGGATCGGAGAAC 16135
QY 212 ----- 212
Db 16134 GTTTCAAATCTAGCCATCAGGTTAGCGTGCATGGGTTACCATGATCGGTTGAATCTT 16075
QY 212 ----- 212
Db 16074 GGTGAGCAATCCATCTGTGATTTTAATTTCCGGGGTGGGAAATGTAGAAAAAGCA 16015
QY 212 ----- 212
Db 16014 GAGGATATTATGAACAATCTGCTATTTATTAATCAATGTCATCAATCATCTCTT 15955
QY 212 ----- 212
Db 15954 GAGAGAGTCTGTATTGATGTTTACTGATAATGTTTGGATCTTTTCACTCTCTTTT 15895
QY 213 ----- 213
Db 15894 AGCTAAGAGTTCTTACCTGAGTCTTTTACAGTTTTCAGAGCTGGCAGCGGCTCCAC 15835
QY 224 apProGlyAlaIleIleIleIleValThrTrpCysHisArgAspLeuGlyProAspLysLys 244

```

```

Db 15834 TCCTGGGCGGAGAAATTAATCATTTGACCTGCTGCATGAGAACTCGAGCATCGAAGA 15775
QY 244 nSerLeuHisProThrProGlnAspLeuLeuLysLysIleCysAspAlaIleTyrLeuPr 264
Db 15774 GTCCCTGAACCTGTATGATGCTGATCTCTGAAAAGATGATGCGATATATATCTCC 15715
QY 264 cAlaTrpCysSerThrSerAspTyrValLysLeuLeuGlnSerLeuSerLeu----- 281
Db 15714 AGACTGTGCTCTCTCTTGTATGATGATCAAAATTTGCCAGTACGCTCTTGAAGTAA 15655
QY 281 ----- 281
Db 15654 AAACTTTTCATGCTGTGAAGCTGTAAGTGAATTTAACTTGAATGATGTTTGA 15595
QY 281 ----- 281
Db 15594 CATCACTTGGGTACCATCCGATTCGATTCCTCCAGACATCTTCATGCTGTGAC 15535
QY 282 ----- 282
Db 15534 ATGTGAATGTGCCAGATATATAGACAGCTGATTTGTCAGAGAACGTGCCCTTTCTG 15475
QY 297 pProAlaValIleArgSerAlaPheThrTrpLysGlyLeuSerSerLeuSerSerGly 317
Db 15474 GCCTGCGTTATTAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 15415
QY 317 yGln----- 318
Db 15414 TATGATCTTGGCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 15355
QY 319 ----- 319
Db 15354 AACTGAACTAATTTGTTGCTGCTCTTATTTGAAGGCGGAGAAAGCATTAAGAGGTG 15295
QY 324 lAlaLeuAlaMetProLeuMetIleGlnGlyTyrLysLysAspLeuIleLysPheAlaIle 344
Db 15294 CAATGCTGATGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 15235
QY 344 lEthrCysArgLysProGlu 350
Db 15234 TCACCTGTGCAAGCCCGAA 15215

```

```

RESULT 9
LOCUS AP003587 343550 bp DNA linear BCT 28-NOV-2001
DEFINITION Nostoc sp. PCC 7120 DNA, complete genome, section 7/19.
ACCESSION AP003587 BA000019
VERSION AP003587.1 GI:17130808
KEYWORDS
SOURCE Nostoc sp. PCC 7120 DNA.
ORGANISM Nostoc sp. PCC 7120
Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
REFERENCE
1
AUTHORS Kaneo, T., Nakamura, Y., Wolk, C.P., Kuritz, J., Sasamoto, S.,
Watanabe, A., Iriuguchi, M., Ishikawa, A., Kawashima, K., Kimura, T.,
Kishida, Y., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A.,
Nakazaki, N., Shimo, S., Sugimoto, M., Takazawa, M., Yanada, M.,
Yasuda, M. and Tabata, S.
TITLE Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120
JOURNAL DNA Res. 8 (5), 205-213 (2001)
MEDLINE 21595285
REFERENCE 2 (bases 1 to 343550)
AUTHORS Kaneo, T.
DIRECT SUBMISSION Direct Submission
INSTITUTION Submitted (02-MAY-2001) Takakazu Kaneo, Kazusa DNA Research
INSTITUTE, The First Laboratory for Plant Gene Research, Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(F, E-mail: kaneok@kazusa.or.jp,
URL: http://www.kazusa.or.jp/cyanobase/,
Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)
FEATURES
Location/Qualifiers

```

source 1. .343550
/organism="Nostoc sp. PCC 7120"
/db_xref="taxon:103690"
/note="synonym: Anabaena sp. PCC7120"
gene complement(94. .411)
/gene="a111719"
complement(94. .411)
/gene="a111719"
/note="ORF_ID:a111719
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB73418.1"
/db_xref="GI:17130809"
/translation="MKKSAILEAVHETAEADLHKAGLMNÖTTREFEHLCPRIEPL
EPLQIKETRESSQSOAVFARILINISTSVÖKWEIGÖKRPSASIKLLHLVKNRGLNS
VLY"
gene complement(504. .782)
/gene="a511720"
complement(504. .782)
/gene="a511720"
/note="ORF_ID:a511720
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB73419.1"
/db_xref="GI:17130810"
/translation="MKIKYKRTFEDMARREGKLNLSLCAVNMENAPGLYDADIGGLF
KRIKAPGCKSGSGFRTLIATNNEVGFLFAÖKTNNAVTLTHMKKRL"
gene 935. .1507
/gene="a1r1721"
935. .1507
/note="ORF_ID:a1r1721"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="BAB73420.1"
/db_xref="GI:17130811"
/translation="MRHKDKKKQAICDAAIELITANGPADTSMKIAKANSPATI
VYFENKEKLINLIVLEVKOEISAETILKGNÖM.SVEKAKMTWNNYKKAIAINPVE
AFTEOPANSPMVRICKDESLSYFKPMLDLFNGCKEKEKVPKDIPIEFTAFETAPILG
LKEHFGSALILDEKMLETTYKIAMAVTN"
gene 1569. .2504
/gene="a1r1722"
1569. .2504
/note="ORF_ID:a1r1722"
probable oxidoreductase"
/codon_start=1
/transl_table=11
/protein_id="BAB73421.1"
/db_xref="GI:17130812"
/translation="MKKEKMAENITIPÖKGRLIAVGTSSGIGETARVILANKOASVI
IAYRNDCKKAKALAKITÖKNDKDVKMEIDLNLASVKNPAEMFERKNYLRDLINN
AGVAPIPSKTTGCFELÖFCTNHLGHALGÖLEFIISTEGSRIVAVSSAHNMGTI
DPPDLNMEÖRSYAKWKAYGDSKLANLFTYELÖRKLDNGIDTLVTASHPÖWTELO
RTAGGIYKLVINGIYAÖDITMGALPTLRAALEGKGAETGPGNCEMRCGYPIVEEN
ELSKÖDALAKKLWVSEKLTVDVKEPFMKKÖSAGK"
gene complement(2701. .3462)
/gene="a111723"
complement(2701. .3462)
/gene="a111723"
/note="ORF_ID:a111723
probable aldolase"
/codon_start=1
/transl_table=11
/protein_id="BAB73422.1"
/db_xref="GI:17130813"
/translation="MSYFSRPLPPVFNSTIEBERLHRKÖRLAAERLFEGKFGESGIAG
HITRDEFTDHFVWNLGKYFGHIRVSDLVLRNDGVVGDALINÖAFAIHÖIH
EARDVVAHAHSHYGGKAMSLGRILDPLODACAPEYEDHALLDDYGTGVALETSECK
RLAETLGNKKALILRNHCILTVGHVTDEARFWYISLERSQÖQLLAEAGÖPKTIIGHE

gene complement(3764. .5167)
/gene="a111724"
complement(3764. .5167)
/gene="a111724"
/note="ORF_ID:a111724
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB73423.1"
/db_xref="GI:17130814"
/translation="MSTÖSGVPGCLDPLTHPELILÖSGÖLITASLLVFLVYAGIA
IALMSFALRHNSÖQRTIFIGEMAIRYÖILRGÖHLATLVLILVGGFLLSTLSNRH
HMEÖAKIAÖVÖVSGDKLEÖIAPÖVRYTÖEPTYNÖNKIVKÖVNDTÖÖVSRLLS
LGSÖIÖVLEÖSTDVÖGRSAVYRÖDYTÖVKVNRKIDINNEFEIÖPÖNGÖSLLSN
KYERÖRTLEPÖNPGNYSRPAFLPEPÖGTFSEFVYVÖAÖGAPRÖVYNSAÖGLSNFRLL
TYIANPGADFASGIVYKREKADGRSTÖYTHIFEDDVÖSVKNPÖGVTNTNPIRÖGTIT
PRLLLAPALELWMLILLYLSPLISLTVNAISITIFACILITLTYISRMÖRÖLANTM
ISLVLLTWGLÖSTRSASILAITCYIAGAVVVLGLLVPSGLTSLVAGLSTVILA
VRHVMNINWÖPÖDSKS"
gene complement(5492. .5902)
/gene="a111725"
complement(5492. .5902)
/gene="a111725"
/note="ORF_ID:a111725
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB73424.1"
/db_xref="GI:17130815"
/translation="MFLKPPSGLMVYASLTLNGLAITYSGRSIAEAKSNFYÖNGA
DÖRGFPALDASLNGTNTFTLVÖDHGIGIAKRRILHAÖSGRLEFTRKFSLYANGKILT
SEDKTEÖEIPFPKPTPDASMEIYCVANNHÖKK"
gene 6321. .6842
/gene="a1r1726"
6321. .6842
/note="ORF_ID:a1r1726"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAB73425.1"
/db_xref="GI:17130816"
/translation="MKSÖHLÖQTATLAGEVSÖQYLKEERTALNATLAELODFIDMCPDARE
VRKALAKIVÖGYLDEIÖKILDSGÖSTIGMÖQVAEÖGIDÖLRINÖGRKSYLNA
KÖDÖEVLÖMÖTKÖMELGELÖYLAIEYDVYÖSKSYIDLEÖEACISMKTKYVP
KADENAVAEKRD"
gene 6919. .7431
/gene="a1r1727"
6919. .7431
/note="ORF_ID:a1r1727"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAB73426.1"
/db_xref="GI:17130817"
/translation="MÖGDLSSYVWÖKSÖEIAVYVNDKÖTYVGAÖVYLLGELVYK
AYDÖGNSANTNYLEYLIANSPDÖRLIFWÖGASIHNSKEIRGELDSVNSÖLPTBÖWK
IHCVAFNCPVÖNPIEDIWÖLÖATWVRFCALIPFSHLKWIEMFIRHTTÖDÖTL
ÖMÖGYÖSKIK"
gene 8064. .9401
/gene="a1r1728"
8064. .9401
/note="ORF_ID:a1r1728
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB73427.1"
/db_xref="GI:17130818"
/translation="MTÖGNARNRINAFRLRVSGNGAFAÖGSESRYSITSKEVYIGRPSÖCO
VLDDMNYRMVSRRAVVRPYVASSVDSKFSWVLDLNSANGTYINGÖRLGÖCÖELHAG

Alignment Scores:	
Pred. No.:	1,77e+45
Score:	659.00
Percent Similarity:	64.29%
Best Local Similarity:	49.29%
Query Match:	36.01%
DB:	1
Length:	343550
Matches:	138
Conservative:	62
Mismatches:	92
Indels:	8
Gaps:	3

QY	245	Series: HSP70, PGP, Lys, Asp, Leu, Ile, Val, Tyr, Ser, Thr, Phe, Leu, Pro, 264
Db	95302	CCACTGACGGCTGATGAAAAAACACCTTAAGACATATTTATCGGGTCTATTGTGCTT 95361
QY	265	AlaTrpCysSerThrSerAspTyrValLysLeuLeuGlnSerLeuSerLeuGlnAspIle 284
Db	95362	TATGTAATTGCTGGTCCGGCATATGACCGATTCGACCTCACTAATCACTTAATAATATTC 95421
QY	285	LysSerGlnAspTTPSerArgPheValAlaProPheTrpProAlaValIleArgSerAla 304
Db	95422	CGCACCGCCGACTGCTGCCAATCCGTCCGCCCAATTTTGGAACATAGTCATCGATTTCGCC 95481
QY	305	PheThrTrpLysGlySerSerLeuLeuSerSerArgLysGlnLysThrIleLysGlyAla 324
Db	95482	TTTACCCCCCAAGCAATATTTGGCTTACTCCGCGAGGTGGACTACCATCCAGAGACC 95541
QY	325	LeuAlaMetProLeuMetIleGlnLysTyrLysLysAspLeuIleLysPheAlaIleIle 344
Db	95542	TTATGACATAGGCTTATATCGTCGCGGCTATGACGCCGGGTTAATATTCGGTTGGCTT 95601
RESULT	10	
SYCSLRP		
LOCUS		
DEFINITION		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
REFERENCE		
AUTHORS		
FEATURES		
source		
gene		
CDS		

gene
986..1933
/gene="slr0073"
CDS
986..1933
/gene="slr0073"
/note="ORF_ID:slr0073"
/codon_start=1
/transl_table=11
/product="sensory transduction histidine kinase"
/protein_id="BAA10540.1"
/db_xref="GI:1001703"
/translation="MGEMPSSKEGTTLDQKILGFIIEAOGHELTLENGCILDCKV
VDSSEVYNEFRASHIKGGAALGTTSTOKTAHRLDEDAKVLKEHQLPVDQKLESLF
LNGYVDLQDLVERLQSPAGIQPEADSIYDGAAYOEELQNYLNLNOSARATSTQ
TMAKSTGTGTENSDVPEIKVLHKMLQLEKQEPSATSRKLEQECYCTGRKLRQ
ENMHLSTLSKQAIANPLHSYRTLAIPVLRKLDKQGYDCFLAKPAGNIRVSDGLQKAS
SOVAOYLIPADPAASVLMKVPNDSQLKQLKNSAAR"
complement(1930..2652)
/gene="sl10088"
complement(1930..2652)
/gene="sl10088"
/note="ORF_ID:sl10088
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAA10541.1"
/db_xref="GI:1001704"
/translation="MGVLSVYWTATHRPQNFLLLTMTLSSSHSTKEDIICYLKKE
GGSAIAMAEEIGISPOAMRKHKDLENGLIEHOEROSMGRRPQLYQSKGRQF
RQRGEFALSTDSLVEYVGEEOGLAVLKQOMRKKEATRQIGGCPLSRYKRLVEL
RKQEGMAEIHPLSVEQAEKFIJSEHHCALADVAESYPTVCGHELMFAALIPDCAIE
RTHMLDNGEHTGYLIQSKTPN"
2908..4350
/gene="ycf24"
2908..4350
/note="ORF_ID:slr0074"
/codon_start=1
/transl_table=11
/product="ABC transporter subunit"
/protein_id="BAA10542.1"
/db_xref="GI:1001705"
/translation="MSTTVKLVNOPYYYGYVTNIEADAIIPGLSEDEVRLISAKKN
EPEFMDELRLRAYRHWLTMAEPTPAVHPPIYDODIIVYSAPKSKKLESIDEVDP
ALLETTEKGIPTSEQKRLSNVAIDAFDSVSGTTEKKEIAEDGVIPTCSISEALQEH
PLVQKYLGSVYPTADNFPALNSAVESDGSEFYFIRPKVCPMELSTYRINNGDGH
PFRLLIIAEGASVSYLECTAPMTVDNOLHAAYVELVALDADIKSYVQNMAYADE
NKGCIYVPTVRGCLCKYNSKISWTQVETGSAITWKYPCVLYCNGNSYGERFSIALT
NNKQADITGTMHIGKNTRSIIITSGISAGNSANSYRGLVKNQKPAQAGARNYSQDS
MLIGDPAANTPEYIQVDNNTAKVEHEASTSKIGEDQLFYFAQRGISBEDAVSMLVSG
PCKDVLNELPMPEFAAEADKLSIKLEGTVG"
4492..5262
/gene="ycf16"
4492..5262
/note="ORF_ID:slr0075"
/codon_start=1
/transl_table=11
/product="ABC transporter subunit"
/protein_id="BAA10543.1"
/db_xref="GI:1001706"
/translation="MSQTLISIKNLASVADGNOILKGVNLEIKACEVHAIMGKNGSGK
STLSVITGHPYEITGGEIITIGODLSALEPHERALAGIFLAFOYPLSENGSNDIF
LRILAYNAKKRKHGLDELODFEDELIOEKLDDVYKMPAPFLERSLNGEFGSGEKKRMEI
LOMALLEPTLSLIDELDSGLDIDALRIYSEGVNPLKNPNMNLVITTHYORLLNLIIPD
HIVHATDKIVYSGGKELALEEKGYDPLDQVLAPI"
5398..6759
/gene="slr0076"
5398..6759
/note="ORF_ID:slr0076
hypothetical protein"
/codon_start=1
/transl_table=11

gene
/protein_id="BAA10544.1"
/db_xref="GI:1001707"
/translation="MTAAILADLQKVHQTETIDKELOAKEDAMVSGTSDSLVYKR
OGAADLGSRLRPHKDEBEQFTDSEKAIIDVYAKGKSLDVAANEFTLPEAHOSR
LVFINGEFTPELSNTNDLPBSAITQSQMNTAAHQREQLAVLQKTDGNGFNSNLTA
GMTDSAVVMIPANTELKSPIHLLEFLTVYDPTLMVOPRLVYVYENNAQVYIAGSYGAI
STNCIDRPOOQPYFNINIVSEIYLGENAQVYHINRORSGSGFHATATYIAQKOSRR
LIDVNGAKLSRNLMOTQOGEATKTEFTALITLLAGOSYDSTHAIANPHGATNKL
HKCIVDEYQAQVRSKGVNDPQAOLNNAOLNNRNLVLSARINTREPELOITADNKC
SHGATTSOLEADEVFYLRSGNDYDARRHLIDAFAGEIIDQPLASVQRLRCVSC
RTI"
6827..8089
/gene="nifs"
6827..8089
/gene="nifs"
/note="ORF_ID:slr0077"
/codon_start=1
/transl_table=11
/product="Nifs protein"
/protein_id="BAA10545.1"
/db_xref="GI:1001708"
/translation="WVALQIPSLAATVRODFPILNDEINCHPLVYLDNANTSOXPRAY
LEKLMHYENDNANVARGAHLQSVRATDAVEARNVYAKFINARSPEIYVTRNATEA
INLVAYSMGMMNLKAGDEIITTVMEHHSNLVPQMYAAKGAVLKFTVQLEOSESFLE
HEFTLSEKTLVTVYHISNTLGCVPABEILADLAHOAGKVLVDQOQSAAPHYLDVQ
LIDCDMLVAGHKMCAPTGIGPLYGKEEILAMPPEFGCGEMIAEVPDFFTGELPH
KREAGPATAEALALGAADYITDLMENHNYEVELTHLMQGLQITPOLRILGPNP
KHDDRALASFVYAGLHSDVATVMDGDAIRSGHHCYQPLRLRFDASGSARALYF
YNTKEIDLEFLOSQATIRFSDDEFTV"
complement(8079..9269)
/gene="sl10086"
complement(8079..9269)
/gene="sl10086"
/note="ORF_ID:sl10086
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAA10546.1"
/db_xref="GI:1001709"
/translation="MRVILMTGKGVGKTSVVAATGIRCAELCHKTLVSTDPAHSLA
DSFDLELGEHPRLVKENLWGAELDALMELEGNMGAVKRYITQVLRAGLDGVAEELA
ILPGMDEIFGLVPMKRRHYDEADYDLIIDSAPYGTALRLSLPEVGGWYRRFYKPIQ
GMSVALRPLEPFRPIAGFSLPDKVMDAPYEFYQIEALEKVLTDNDOTSVRLYN
PEKMWIKESLRAHYSLVNSYDVLIANRILPETIDDPFORWKNOOYVYKQETIVDN
FHLRIPKEAKSEEMCGLAELERLMDTKLXKDDPSQYIKENTINIVQSSNDYSIE
LVLPKPKQIQLNKGTGDELVNRTIGNHRRNLVLPQALALSPAGAMMEDYLRIRAE
AVKR"
complement(9312..10523)
/gene="sl10085"
complement(9312..10523)
/gene="sl10085"
/note="ORF_ID:sl10085
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAA10547.1"
/db_xref="GI:1001710"
/translation="MNSDRSRODPLFISLIPHLMGKGHIVPYHGAVRSVEYLQMY
QAMVPQDADPNLEIEDRLRILPEDLAEESVAKITRLPAMQCSVDYLAQOQLSL
LNGESRPVITLERFIHLQFLAIIWMLQRLSEARSRLHFWILYRDDVLDQTSIY
HMLHORLRLKLPPOCOPLTDSPPGLRSLODYGGPFTVPIHTPOFAEVLPFTEDR
PLTCMPPGAPREKGGKSVMRHYSILPRPSPLQLIAAASGLKSSGSEVNLIGDR
LITREBAGLAKNAQIILILPYDEPVAERTSGITTEAIIAERTPLVSPGTMMARELMQY
DIDSLVDNQAOOSLVVAEITDKLADSSIHQOLAAMAQAVRRHHSLESYAKTIOETWLS
RANGNNEVNMN"
complement(10520..11005)
/gene="hisB"
complement(10520..11005)
/gene="hisB"
/note="ORF_ID:sl10084"
/codon_start=1
/transl_table=11
/product="imidazoleglycerol-phosphate dehydratase:"


```

/gene="SB234M12.3"
/note="highly similar to Oryza sativa gene for putative
receptor kinase GB:AF128457 GI:4680335; similar to Oryza
sativa receptor kinase-like protein gene GB:AF119222
GI:4680488; similar to Arabidopsis thaliana receptor-like
protein kinase emb:AL138654 GI:689879"
/evidence=not_experimental
complement(join(<12094..12725,12823..14263))
/gene="SB234M12.3"
/product="putative receptor kinase"
complement(join(12094..12725,12823..14263))
/gene="SB234M12.3"
/codon_start=1
/evidence=not_experimental
/product="putative receptor kinase"
/protein_id="AAM94320.1"
/db_xref="GI:22208505"
/translation="MYTLAPRLPRLFLLLAAAAAANGNNTTEEAAMALVNL
KSEPTDPTGRLKMSATSPFPCDASPWFQVOCYKSLVGRLTNNLSGTDEFGAI
AKLRLHSVNLKHNALSGPLPASLGLRLGLALYLSNNFSGPIPADVFANMLKTL
YLDNNRITGCLPADAIKAPRLMELHLDNRQIDPPVFKLPESLRNVSHNRSLGTI
PQSAERYDASFPAGNGLGSCPSDAVCAAGPALPAPPTADYAMEESTSV
FVYIGITLIVLLSGAMVLMRLDDENSVATAMDYAGSTTGAGAGAGSGSKSA
DAASTAARAGEPAVAVDMASGSSSSHGGGSSSRBMGEYLLMEDIPATGLPDL
MKASAYLNGTGSATKAAARMGVTAVARLRDMNRFGRDEEQHVHMGQLRHPV
LPVGHTRKKEKLIYSEYPRCSLYLHGDSPNVLIDMGRVRIAGVVGALF
LHRLGIPARLVSMDGADDPAPPDGNLGSNILLGADMEPLLDVGFPLVA
AQAPQAMFARSEGTGRGVASDVYCGVYLLETGTFRPSOYLISRGCDVYH
MAATVAAGEAEVLDPALIAAGDAAVRLRLRGVHCAPEPECRPSVALAAMVERI
GSAAGGAS"
12558..12578
/rpt_unit=c99
complement(13145..13307)
/rpt_unit=c99
13490..13517
/note="GC-rich"
14190..14216
/rpt_unit=c99
complement(<15173..>20442)
/gene="SB234M12.4"
/note="similar to Arabidopsis thaliana gene for putative
sugar transporter protein GB:AY094465 GI:20453188"
/evidence=not_experimental
complement(join(<15173..15605,16042..16226,16382..17055,
17854..18423,18531..18650,19447..19501,19596..19649,
20293..>20442))
/gene="SB234M12.4"
/product="putative sugar transporter"
complement(join(15173..15605,16042..16226,16382..17055,
17854..18423,18531..18650,19447..19501,19596..19649,
20293..>20442))
/gene="SB234M12.4"
/note="ATG35300/F23E12_140"
/codon_start=1
/evidence=not_experimental
/product="putative sugar transporter"
/protein_id="AAM94321.1"
/db_xref="GI:22208506"
/translation="MSMICKLNVQAEDLMGATGAGQGVOKNGCKLEVAVLGCAL
RIMAAVSTYHOVCAMTFRFACATASRLIDELATREFSGFIIVCKRAIVLDMK
STVSAVVVSTIGAILMDPFALEENHMKKEFLNGSIEGTLAASTFGAIVIT
IFSGLLDMLGRRAILVSSLVSGVLMSPNIVYLARILVSGSGVATFVP
IYISETSPNNRGLGTMPQMFICTIFSICLIPWILKSSNNHMGAIIPASIV
YFALVYVLPESPRLASDGKISEARVSLORLGRKHDSGETAIVDCVDTISDAV
GTARASFSGTASRTWSPSTPYWOLSDPLVDLGSIHENMSGGSRNSFFVFNSF
SPREHNNHNRDGNDOOTREATISAGEVANNGLASLISQAASVEVDANSTPSE
GSSVTLRRHGTSLAQEFMASTHDHDEEEELHGFVLPHQSAAPRMESGRHOVRL
SETADSKFMKRVLLQPIRHALCYGMILALQOSTISGLRCAPILREOVGSLSD
IGLSPHSTSLISTLQALLPCTITAMLMDCGRVIALTPILLISLWSMSI
LFNMGPDAIYHFAITVOCFSYVIGLCPPIPIILSEIPTNARATCFSCLSWF
GGLLSAYCPLPVMKSTIGLGCAGCIYALVCCAPLFLFYIRIPETKMLNELIAELFRLS
ROEYV"
complement(22125..22161)

```

```

/repeat_region
/note="AT-rich"
complement(22424..22446)
/repeat_region
/note="AT-rich"
23019..23089
/repeat_region
/note="AT-rich"
23130..23172
/rpt_unit=ta
23495..23526
/note="AT-rich"
26812..26900
/repeat_region
/note="AT-rich"
37014..37038
/note="GC-rich"
complement(37399..37462)
/rpt_unit=a
<37578..>44288
/gene="SB234M12.6"
/note="similar to Nicotiana plumbaginifolia ubp1 gene
emb:A172011 GI:6996559"
/evidence=not_experimental
join(<37578..37721,38745..38806,38998..39100,40241..40317,
40518..40579,40690..40777,40913..40988,41080..41110,
41205..41318,42137..42222,42364..42537,43066..43258,
43335..43537,43864..43887,43980..>44288)
/gene="SB234M12.6"
/product="putative oligouridyate binding protein"
join(37578..37721,38745..38806,38998..39100,40241..40317,
40518..40579,40690..40777,40913..40988,41080..41110,
41205..41318,42137..42222,42364..42537,43066..43258,
43335..43537,43864..43887,43980..44288)
/gene="SB234M12.6"
/codon_start=1
/evidence=not_experimental
/product="putative oligouridyate binding protein"
/protein_id="AAM94322.1"
/db_xref="GI:22208507"
/translation="MHPQRMKQAADQOLMQALLQOOAOOPLEPGHHHPGLL
AAROMEPVSGNLPFGPDSSTGRSVYGNISLQTDVIVLOEVSIGPVGCKLIRE
KSSRFQTDYDRIYAAIALISLNGRPLQGPITVYNNATSTQEDISGHNITVDIC
PEYTDATLPAFSGSITCSDPARWMDKTRSGFGVSTRNOODQSAINDINGKWL
GNQOIRCNMATKGANAGEEKOIYDSKYDLNIGTSESGKNPDGFSNPOFTTVYG
NLPEHATNNDVHLFPHSLGAGSIEEYAVTRDKGFGFVRYSTHEAQAQIOMANGOLIG
GROIKTTIYNGVYKLRVAEESDFCEVLYLSRLVTVAASLVAVFLLEGDAYRO
GIDLLAASVAGAVGCMGVCGLVLTLYVAOSPDGILLTMKAGACACIAYIR
SKLGGSVKCAKCAKGAAYVAVTBNINAKSKSKSPPTTASSPLPPAPALFSTGV
SPDPLAIONMALINKATNPISLMGQHALNOVALMGGRASQSTYDGGFQGISMAATT
SAQQOQOQOQOLMY"
37580..37668
/rpt_unit=cag
complement(39121..39142)
/note="AT-rich"
44247..44274
/rpt_unit=cag
complement(44252..45289)
/rpt_unit=c99
<45290..>46757
/gene="SB234M12.7"

```

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
4.17e-39	108553	579.00	160	38.908	61	242
Percent Similarity:	38.908					
Best Local Similarity:	32.598					
Query Match:	31.648					
DB:	8					

US-09-857-613a-28 (1-350) x AF527809 (1-108553)

QY 94 AspSerAspSerThrValSerLeuSerAspHisArgAlaIaIaGlnIleArgMetIleGln 113

DB 94288 AATTACACTAGTATATATAAATTGGTTGAC-----ATCGTCCCACTTGA 94241

QY 114 GluSerLeuArgPheAlaSerValSerGluGluArgSerLysTrpProLysSerIleVal 133

Db	94240	CTAGATTTCATGCGTATTTTCTGTGCGACATGATCATAGAGAGAGGCCCAAAACATATGCTC	94181
Qy	134	AspValGlyCysGlyIleGlyIleSerArgIleLeuAlaLysLysPheGlyAlaThr	153
Db	94180	GACGTAGCATGTGGCATTTGGAGGCGAGCTCAAGTACTTGGCCAGAAAATACGGAGCACAA	94121
Qy	154	SerValGlyIleThrLeuSerProValGlnIleIleArgAlaAsnAlaLeuAlaAla	173
Db	94120	TGTACGGCATACCTTGAGCCCTGTTCAGCTGAGAGGAAATGCTCTTACTGCGGCC	94061
Qy	174	GlnGlyLeuAlaAspLys-----	179
Db	94060	CAGGCTGTGTCAGACCA-GGTGCCAAACACTTCCGCTCCCTTAATTATTCGTGATTATG	94002
Qy	179	-----	179
Db	94001	TACTATTATTTATCATCATTCATAGTAAGTACGCGCTAAGAGAGACTTGTCAATATCA	93942
Qy	179	-----	179
Db	93941	GTCCGTCTCTTCACACACCTGCTTTTCAGTTTTCATCGAGATCATGTAACGATGCTGATG	93882
Qy	180	---ValSerPheGlnIleValAlaAspAlaLeuGlnGlnIleProPheSerAspGlyGlnPheAsp	198
Db	93881	CAGGTTTCCCTCCCAAGTTGCTGATGCTCTGAGAACACCTTCCGTGATGAGATGTTGAT	93822
Qy	199	LeuValITripSerMetGlnSerGlyGlnHisMetProAspLysAlaLys-----	214
Db	93821	CTTGTATGATGTCATGATGAGAGTGTGTAGACATGCCGATAAAG-AAAGGTAACTGTTT	93763
Qy	214	-----	214
Db	93762	TTTTCTGTGTTTGTGCTTGAACACTCTTAACCTCCAGAAAATGACCAAAATTTCTTGA	93703
Qy	214	-----	214
Db	93702	CTTCGTTGCAAGTAAACCTGATATGTTTGAACCTTACCTGATTTGTATTAGTTTAG	93643
Qy	215	-----PheValGly	217
Db	93642	TGGCGTAACGTGATATGATGTGACACTTACCTTAATTACTTGTGTATCAGTTTGTAAT	93583
Qy	218	GluLeuAlaIleArgValAlaAlaProGlyAlaIleIleIleIleValThrTProPheHisArg	237
Db	93582	GAGCTAGCCGCTGCTGCACCTCTCTGGAGGACAAATATCATGCTGCATCAGTCCCATAGG	93523
Qy	238	AspLeuGlyProAspGlnIleSerLeuHisProTPrGlnIleAspLeuLysLysIle	257
Db	93522	AACCTCAATCCGCTGAAACCTTCCGTAAGAACCCGATGAACGAGACTCTTAAGAGGATA	93466
Qy	258	CysAspAlaIleTyrIleuProAlaITrPcysSerThrSerAspTyrValLysLeuGln	277
Db	93462	TCTGATGACATACGTCTTGTCCACAGCTGTTCTTACCTTCAGACATGTACGATTTGCTAAG	93403
Qy	278	SerLeuSerLeu-----	281
Db	93402	TCGACTGTCTCT-CGAGGTAAACAAAAAAAGAAAGAGTGTTCGTATGTGATCTGA	93344
Qy	281	-----	281
Db	93343	TAAITCTGAATGAATACGAAAGACCTTGTATGCTTACCTTGCAATTCGAAATACGCTT	93282
Qy	281	-----	281
Db	93283	GATCATATCTTGATGATAGAATTGAATGCTTCACTTACTACTAGGAGCACAGCCGCAAT	93224
Qy	281	-----	281
Db	93223	TTGGAATTTGTTTGCATGATGTTAATTTTCATAGAAAACAATCTATTTTGTTCGATG	93164
Qy	282	-----GlnAspIleLysSerGlnAspITripSerArgPheValAlaProPheITripProAla	299

[illegible]

```

gene
CDS
    complement(1482..2213)
    /gene="ho2"
    /complement(1482..2213)
    /gene="ho2"
    /note="ORF_ID: t111721"
    /codon_start=1
    /transl_table=11
    /product="heme oxygenase 2"
    /protein_id="BAC09273.1"
    /db_xref="GI:22295447"
    /translation="MTVLSLARREGTAEAHTLAHTAFMCEVRGLITSRLRPROFLAN
LYEVYDLEALASANSDAVAIYFPELINSQGLAODLAEFYGEDMRDQVPTLATRY
YLSRLHDVAOSEPLLLLAHSYTRYMGDLSGOALSKIVHSOLMLPEKGTAFFEPAL
PTMEDKAFKORYRATLDSGLDEALARIYQEAENFAPALNCNMHMLSESLKTTICD
QVMOSWIENQPRROHPRPEALAI"
    2486..3550
gene
CDS
    2486..3550
    /gene="t111722"
    /note="ORF_ID: t111722"
    /hypothetical_protein="
    /codon_start=1
    /transl_table=11
    /protein_id="BAC09274.1"
    /db_xref="GI:22295448"
    /translation="MVAIAPNPSGSGSVKTIKREMLTLPREYTDPEQAOKLIDSRQOE
OLEAMLEMRADYNRRHFRVDRDPSFKKIDLDNOTKRAFDIYERSCVSEPSGTLRK
ELSRKLDKRRPLAEIHFHLMARDBARHAGLNKAMDFGAMDGYSKRTYTFEPL
PWLTVLYSEKIGVRYIILYHLEKHPEHSFYPLFNYFERMCOENRIGDIFKALIL
HSQPOLIOGAKLMRNFLLTVAFHTTLVHERAKVEALGDAPADEVIAKTNE
TAARFESVLANVDHPEFSYRLQRCVEISNQLQALEATHQKMKALKRKLPHQALIAH
LRLTYLLPVNAQOEWGTVH"
    complement(3547..3858)
    /gene="t111723"
    /complement(3547..3858)
    /note="ORF_ID: t111723"
    /hypothetical_protein="
    /codon_start=1
    /transl_table=11
    /protein_id="BAC09275.1"
    /db_xref="GI:22295449"
    /translation="MLSRPLIERLAEIGRDYIMDAKWHLYIGDAKLATPLAEAFI
PLLEGGEMSLAQVTSVLQEVSVPLGGDQQLSLERLIPKANQELVLELNRFOEELS
S"
    complement(3963..4382)
    /gene="psad"
    /complement(3963..4382)
    /gene="psad"
    /note="ORF_ID: t111724"
    /codon_start=1
    /transl_table=11
    /product="photosystem I subunit II"
    /protein_id="BAC09276.1"
    /db_xref="GI:22295450"
    /translation="MTTLTGQPLVYGGSGGILISADTEEKYAITMSPKEQYFEMPT
ACAAVMRBCENLVYPRARKEOCLALAQQLRPKRINDYKTRIRPDEGTVLHPRKQGV
PEKVNKGRVANSVPRSIGQNPNSQLKFTGKKPYD"
    4763..5398
gene
CDS
    4763..5398
    /gene="t111725"
    /note="ORF_ID: t111725"
    /hypothetical_protein="
    /codon_start=1
    /transl_table=11
    /protein_id="BAC09277.1"
    /db_xref="GI:22295451"
    /translation="MOYLYLHGFASSPRSSKAOYFFDRREBELCQPLLIIDLNQDEFFH
LTLRQIDQVEALLPPEPITLLIGSSFGGLTAAWLAEKHPQVQVLEFLAFAFEFAIHM

```

```

gene
CDS
    complement(5383..6369)
    /gene="t111726"
    /complement(5383..6369)
    /gene="t111726"
    /note="ORF_ID: t111726"
    /codon_start=1
    /transl_table=11
    /product="delta(24)-sterol C-methyltransferase"
    /protein_id="BAC09278.1"
    /db_xref="GI:22295452"
    /translation="MSHGLITVTVVALVLLGLALYLEPRKYESARSVAESYDMNT
KDGLIEFWGCHIHGHYGLPPRPKDFROAKVDPEHVMVRAGLDRLPCTTYIVDVC
GIGSSRLRAGDVEHVTGITSPEQYRRARELTPRALNRFQDDMLASFPSASD
VVMSEKGRPHRPMQOQAKELLIRLKGCGITLVADMRORDROOPLNWEELIROLL
DOMHAPARASIEGFBEALATGLVAGVMTADMTQETLPSLDSIWOGIVAPCEGLIRF
GLPLVKSILREVPFLLMRIAFGMGLCRFGMFRAVRAEIRAVLESPAPQVNC"
    6619..8262
gene
CDS
    6619..8262
    /gene="riba"
    /note="ORF_ID: t111727"
    /codon_start=1
    /transl_table=11
    /product="riboflavin biosynthesis protein Riba includes
GTP cyclohydrolase II and 3,4-dihydroxy-2-butanone
4-phosphate (DHBP) synthase"
    /protein_id="BAC09279.1"
    /db_xref="GI:22295453"
    /translation="MPVFDSTIESALALRGEVIVVVDENRENEGDIGAEERVPA
MINMFAVARLILICLAMEGDDELDELPLWTTNTDSNOTAFYSVDAGRMVCTTGI
SAERPARTLIDPSPODLPREGHVPFLRSRGVILRAGTEAAVDLTRLAGLY
PAGYICIQNDGMSRLPOLMEARTHQIKTISADLSTRQHEFRVREVAKLIP
TERGEFOITGRNSLDQSEHAIVKGDPAFGDPVAVRHSCTGDAIGSLRCDOR
MQDLAALKMIINAGAGVYVLRQEGRIQVNRKRAYSLQDLMDVEANENHGFAD
LRNYGVAQIILNDIGVKQIRLITNNPKIAGLKGLEVDVRYPLILEATPYNTPLYT
TKAEKLGHLILQTYLITVAEFMWSGOGLAADREYERLEKHLHLAGVHLVREEARPA
QALRGHDLVYHGEFDQPHVADRQWYQOPHPVYLAVQTLIRFOCCMPLTKGFEFLVA
TGDDPMNLPMDDLRQAYTQOOSPSBMRPHLITYMSAAG"
    complement(8267..8453)
    /gene="ts11728"
    /complement(8267..8453)
    /note="ts11728"
    /note="ORF_ID: ts11728"
    /unknown_protein="
    /codon_start=1
    /transl_table=11
    /protein_id="BAC09280.1"
    /db_xref="GI:22295454"
    /translation="MDVRRDKQCHNSYISVIYWKSLAPLKKRVTTRPEKGVNLGVVP
TPQSQTCCQTSQTSIGAF"
    complement(8456..9043)
    /gene="t111729"
    /complement(8456..9043)
    /gene="t111729"
    /note="ORF_ID: t111729"
    /similar_to_herit_protein="
    /codon_start=1
    /transl_table=11
    /protein_id="BAC09281.1"
    /db_xref="GI:22295455"
    /translation="MMDCPPLPIYVQMRAPHRNLTIDSSALLKMLTLPDPVPROVI
LRAYLRKTOPNDEBRLPRAAGKPTLNGLEFMSHSGNLAVLAVSGRAVGVDEIL
RLCQPARAISRRFEGALQORILEGGDRSLQAMTYEAWLKAGIGVQWRRAOAGLE
HMVASPEVGDRAISVVLTPPPQCFEFPPEMAA"
    complement(9050..9559)
    /gene="t111730"
    /complement(9050..9559)
    /gene="t111730"
    /note="ORF_ID: t111730"
    /hypothetical_protein="
    /codon_start=1

```


QY 87 Hismethishisgilypheryraspsr-----AspserthrValSerleu 101
 Db 127439 CATATCCATTGATGATCATTTATGCTGCACCTCAAGAAAGATTTCGTGGCTAA 127380
 QY 102 SeraphisargalalaglinleargmetileginserserleuaphelaserVal 121
 Db 127379 TCTGAT-----TTTGTCCATGAATGCTCGTTGGGCTGTTG 127341
 QY 122 SerugluarserlystrPro-----LysSerileValaspValglyCysglyle 139
 Db 127340 -----GATTAACACCCCTGCTACTACCTGTTGATGATGTTGGTTGGATNT 127293
 QY 140 GlylyserSerarglyLeuAlaLysLysphelyalathrservalglylethrleu 159
 Db 127292 GGGGGTACTCCATTTTGGACAGGATTAATGATTGGCTTACAGATACCATC 127233
 QY 160 SerProValGlnAlaLysLysLysLysLysLysLysLysLysLysLysLysLys 179
 Db 127232 AGCCCCACCAAGTCCAAAGCGCTCAAGATTACACACAGAACTGAATGCA----- 127179
 QY 180 ValSerPheGlnValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 199
 Db 127178 ---CAGTTTGGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 127122
 QY 200 ValTrpSerMetGlnSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 219
 Db 127121 GTTGGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 127062
 QY 220 AlaArgValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 239
 Db 127061 ATCCGGTACTAAAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 127005
 QY 240 GlyProAspGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 258
 Db 127004 -----GATGCCAAACCCCAATTTTGGGAGAAACAGTAATGACAGCAACTCTA 126951
 QY 259 AspAlaIleThrLeuProAlaIleThrLeuProAlaIleThrLeuProAlaIleThr 278
 Db 126950 GATCAGTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 126891
 QY 279 LeuSerLeu-----GlnAspIleLysSerGlnAspIleLysSerGlnAspIleLys 296
 Db 126890 ACCGGATAGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 126831
 QY 297 TrpProAlaValIleArgSerAlaPheThrTrpLysGlyLeu-----SerSerLeu 313
 Db 126830 TGGCTTGATTTATC-----TGGCAAGGATAGATTAGACCAGAAAGATT 126786
 QY 314 -----LeuSerSerGlyLysThrIleLysGlyAlaLeuAlaMetProLeu 329
 Db 126785 GTGCGTTTGGTGTATGCTGATTTCAATTAATCTGCGAAGAGTCCCTACTACTAG 126726
 QY 330 MetIleGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 343
 Db 126725 ATGAGCGTGCATGCGTACAGAGCTGTGATTTGGATG 126684
 RESULT 14
 AX073663 930 bp DNA Linear PAT 06-FEB-2001
 LOCUS AX073663
 DEFINITION Sequence 7 from Patent WO0104330.
 VERSION AX073663
 KEYWORDS
 SOURCE
 ORGANISM *Synechocystis* sp. PCC 6803.
 Bacteria; Cyanobacteria; Chroococcales; *Synechocystis*.
 REFERENCE
 AUTHORS Herbers, K., Badur, R., Kunze, I. and Geiger, M.
 TITLE Identification and overexpression of a dna sequence coding for
 2-methyl-6-phosphogluconate-methyltransferase in plants
 JOURNAL Patent: WO 0104330-A 7 18-JAN-2001;
 Sungen GmbH & Co. KGAA (DE)

FEATURES
 source
 CDS
 Location/Qualifiers
 1..930
 /organism="Synechocystis sp. PCC 6803"
 /db_xref="taxon:1148"
 10..915
 /note="unnamed protein product"
 /codon_start=1
 /transl_table=11
 /protein_id="CAC28417.1"
 /db_xref="GI:12710085"
 /translation="MAAGLYLLTARGVSDSVANAYDQMTEDGLEIYRWGDHILGH
 YGDPVAKDFIOSKIDFVHMAOMGIDTLPRTGTVLDVGGIGSSSILKDKGFNV
 TGITSPQVRRARELPDPTAKFADDAVALSPDSFVWSVSGPMPDKAVF
 AKELLRYKREGILIVADWNRDRQVPLNFWERKPYVMQLDQWSPAFASIEGFAEN
 LEATGLGQVTTADWTVPTLPALWLDITWQIIRPOGWLQYGIKRSVREVPITLL
 MRLAEGVLCRFNFKAVRNKATQA"
 BASE COUNT 205 a 230 c 262 g 233 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2.37e-25 Length: 930
 Score: 392.00 Matches: 92
 Percent Similarity: 51.76% Conservative: 40
 Best Local Similarity: 36.08% Mismatches: 87
 Query Match: 21.42% Indels: 36
 DB: 6 Gaps: 10
 US-09-857-613a-28 (1-350) x AX073663 (1-930)
 QY 69 IleAlaGluPheThrYraspGlu-----SerSerLysLeuTrpGluAsnIleTrpGlyAsp 86
 Db 64 GTGGCCAAACCCCTACGCAATGACAGACAGACGCAATTTGGAATATTACTGGCGGAC 123
 QY 87 HismethishisgilypheryraspsrSerSerThrValSerLeuSerAspHisArgAla 106
 Db 124 CATATCCACCTCGGCATTTAT-----GGCATCCGCAATGCGCCAGAT----- 168
 QY 107 AlaGlnIleArgMetIleGlnGlnSerLeuArgPheAlaSerValSerGluGluArgSer 126
 Db 169 -----TTTCATCCAAATCGAAATATTGATTTGTC-----CATGCCATGGCC 207
 QY 127 LysTrp-----ProLysSerIleValAspValGlyCysGly 138
 Db 208 CAGTGGGGGGGATTAATACATCTTCCCGGCAACAGGATATTGATGATGGGTTCGGCC 267
 QY 139 IleGlyGlySerSerArgTrpLeuAlaLysLysPheGlyAlaThrSerValGlyIleThr 158
 Db 268 ATGGCGGTAGCAGTGCATTTCTCGCCAAAGATTATGTTTACGTTACCGCATCAC 327
 QY 159 LeuSerProValGlnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 178
 Db 328 ATTAGTCCCAACAGGTGAACGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 384
 QY 179 LysValSerPheGlnValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 198
 Db 385 -----AATTTGCGGTGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 438
 QY 199 LeuValTrpSerMetGlnSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 218
 Db 439 GTAGTTTGGTGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 498
 QY 219 LeuAlaArgValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 238
 Db 499 TTACTGCGGTGCTGTAACACAGGGGGGCTTGTGGTGGCGGATTTGGAATCAACCGGAC 558
 QY 239 LeuGlyProAspGluGln-----SerLeuHisProTrpGluGlnAspLeuLysLysIle 257
 Db 559 -----GATGCCAAGTCCCTCAACTTCTGGGAAACCAAGATGACGACAACTG 609
 QY 258 CysAspAlaTrpTrpLeuProAlaIleThrSerSerTrpValSerValLysLeuLeu 277
 Db 610 TTGCATCAATGTCACCCCTGCTTTCACAGCATTTGAAGGTTTTCGCGAATAATTGGAA 669

QY 278 SerLeuSerLeuGlnAsp-----IleYsSerGluAspTrpSerArgPheValAlaPro 295
 Db 670 GCCACGGGTTGGTGGAGGCGGAGTGTCTGCTGATTGACGTACCGACCTCCCC 729
 QY 296 PheTrpProAlaValAlaIleArgSerAlaPheTrpLysGlyLeu 310
 Db 730 GCTTGCTTCGATACCATTT-----TGGCAGGGCATT 759
 RESULT 15
 AX073657 957 bp DNA linear PAT 06-FEB-2001
 LOCUS AX073657
 DEFINITION Sequence 1 from Patent WO0104330.
 ACCESSION AX073657
 VERSION AX073657.1 GI:12710078
 KEYWORDS
 SOURCE Synecocystis sp. PCC 6803.
 ORGANISM Bacteria: Cyanobacteria; Chroococcales; Synecocystis.
 REFERENCE 1 (bases 1 to 957)
 AUTHORS Herbers, K., Badur, R., Kunze, I. and Geiger, M.
 TITLE Identification and overexpression of a dna sequence coding for
 2-methyl-6-phytylhydroquinone-methyltransferase in plants
 JOURNAL Patent: WO 0104330-A 1 18-JAN-2001;
 Sungen GmbH & Co. KGaA (DE)
 FEATURES
 source 1..957
 /organism="Synecocystis sp. PCC 6803"
 /db_xref="taxon:1148"
 CDS 1..957
 /note="unnamed protein product"
 /codon_start=1
 /transl_table=11
 /protein_id="CAC28415.1"
 /db_xref="GI:12710079"
 /translation="MPEYLLPAGLISLAIAGLYLITFANGYQSDSVANAYDWT
 EDGILEYWGDIHLGHTGDPVAKDFIDFVHAMOMGIDTLPGTVLVGC
 GIGGSRLIADKYGFNVGTITSPQVKRAETLPDPYAKFAVDAMALSPDGSFD
 VYMSVAGPHMPDKAFKELRYVKGILVADBNODDPOVPLNPEKPYMRDL
 DQMSHPAFASIEGPAENLEATGTVGVTADMTVPTLPMADLTITGIIIRQGMLOY
 GIGFTKSYREVPITLLMRLAFGVGLCRFGMEKAVRNKATQA"
 BASE COUNT 203 a 244 c 268 g 242 t
 ORIGIN
 Alignment Scores: 2.46e-25 length: 957
 Pred. No.: 392.00 Matches: 92
 Score: 51.76% Conservative: 40
 Percent Similarity: 36.08% Mismatches: 87
 Best Local Similarity: 21.42% Indels: 36
 Query Match: 6 Gaps: 10
 DB: US-09-857-613a-28 (1-350) x AX073657 (1-957)
 QY 69 IleAlaGluPheTyrAspLlu-----SerSerGlyLeuTrpGluAsnLleTrpGlyAsp 86
 Db 106 GTGGCCACGCTTACGACCAAGAGACAGGATTTGGATATTAATCTGGGCGAC 165
 QY 87 HisMetHisHisGlyPheTyrAspSerAspSerThrValSerLeuSerAspHisArgAla 106
 Db 166 CATATCCACCTCGCCCATTTAT---GGGATCCGCCAGTGGCCAAGAT 210
 QY 107 AlaGlnIleArgMetIleGlnGluSerLeuArgPheAlaSerValSerGluGluArgSer 126
 Db 211 -----TTCATCCAAATGCAAAATGATTTTCTC-----CATGCCATGCC 249
 QY 127 LysTrp-----ProLysSerIleValAspValGlyCysGly 138
 Db 250 CAGTGGGCGGATTAAGATACCTCCCGCCACACGATATGATGGGTTGGCGC 309
 QY 139 IleGlyGlySerSerArgTyrLeuAlaLysLysPheGlyAlaThrSerValGlyLleThr 158
 Db 310 ATTGGCGGTACAGATCGCATTTCTGCGCAAAAGATTATGCTTTAAAGTTACCGGATCACC 369

QY 159 LeuSerProValGlnAlaGlnArgAlaAsnAlaLeuAlaIleAlaGlnGlyLeuAlaAsp 178
 Db 370 ATTAATCCCAACAGAGTGAACAGGCGGACGAGATTAACCTCCCGATGTACGGCC--- 426
 QY 179 LysValSerPheGlnValAlaAspAlaLeuGlnGlnProPheSerAspGlyLeuPheAsp 198
 Db 427 -----AACTTTGCGGTGACGATGCTATGCTTCTTTCTTGACGAGTGTGTTCCAC 480
 QY 199 LeuValTrpSerMetGluSerGlyGluHisMetProAspLysAlaLysPheValGlyLeu 218
 Db 481 TTAGTTTGTGTGGAGACAGGAGGCGCCACATGCTGACAAAGCTGTGTGGCAAGGA 540
 QY 219 LeuAlaArgValAlaAlaProGlyAlaIleIleIleValIleThrTrpCysHisArgAsp 238
 Db 541 TTACTGCGGCTGTAACACGAGGCGCATTTCTGTGTGGCGGATTTGATACAGGAC 600
 QY 239 LeuGlyProAspGluGln-----SerLeuHisProTrpGluGlnAspLeuLysIle 257
 Db 601 -----GATGCGCAAGTCCCTCACTTCTTGGAAGAAACAGATGACGACACTG 651
 QY 258 CysAspAlaTyrTrpLeuProAlaTrpCysSerThrSerAspTyrValLysLeuGln 277
 Db 652 TTGGATCATATGCTCCACCTGCTGCTTGGCCAGCATTTGAGGTTTGGGAAATTTGGA 711
 QY 278 SerLeuSerLeuGlnAsp-----IleYsSerGluAspTrpSerArgPheValAlaPro 295
 Db 712 GCCACGGGTTGGTGGAGGCGGAGTGTCTGCTGATTGACGTACCGACCTCCCC 771
 QY 296 PheTrpProAlaValAlaIleArgSerAlaPheTrpLysGlyLeu 310
 Db 772 GCTTGCTTCGATACCATTT-----TGGCAGGGCATT 801

Search completed: March 19, 2003, 08:05:54
 Job time : 2726 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 19, 2003, 05:11:11 ; Search time 221 Seconds
(without alignments)
3566.513 Million cell updates/sec

Title: US-09-857-613A-28
Perfect score: 1830
Sequence: 1 MATVRIPTSCIHITFRS.....IEGYKDKLIKFAITCRKPE 350

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-MODEL=frame_plus_p2n -DEV=xlh
-Q=/cgn2_1/USPTO-spool/US09857613/rumat_12032003_100035-6003/app_query.fasta_1.519
-DB=N_Geneseq_101002 -QFWT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum2 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09857613.qcgn_1.1_200_@rnat_12032003_100035-6003 -NCPU=6 -ICPU=3
-NO_XLPRY -NO_MMAR -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV=TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
N_Geneseq_101002: *
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: *
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: *
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: *
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: *
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: *
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: *
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: *
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: *
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: *
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: *
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: *
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: *
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: *
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: *
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: *
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: *
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: *
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: *
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: *
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: *
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: *
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: *
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: *
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1830	100.0	1189	21	AAA29164	Soybean gamma-toco
2	1668.5	91.2	1331	21	AAA29154	Soybean gamma-toco
3	1173	64.1	1257	21	AAA29165	Soybean gamma-toco
4	1162	63.5	1102	21	AAA29161	Corn gamma-tocophe
5	1154.5	63.1	1293	21	AAC50535	Arabidopsis thalia
6	1151.5	62.9	1071	21	AAZ98317	A. thaliana gene 1
7	1151.5	62.9	1350	24	AAL46039	Arabidopsis gamma-t
8	1151.5	62.9	1790	20	AAE17789	Arabidopsis thalia
9	1147.5	62.7	1296	21	AAC34022	Physcomitrella pat
10	1022.5	55.9	1862	22	AAHA4261	Wheat gamma tocoph
11	956	52.2	1011	21	AAA29155	Corn gamma tocophe
12	778	42.5	792	21	AAA29151	Synechocystis gamm
13	647.5	35.4	954	20	AAE17788	Synechocystis PCC6
14	392	21.4	930	22	AAE26173	DNA encoding a met
15	392	21.4	957	22	AAZ61599	Synechocystis PCC6
16	392	21.4	957	22	AAE26169	Synechocystis 2-me
17	392	21.4	957	24	AAL46040	Partial gamma toco
18	392	21.4	974	22	AAE26170	Partial gamma toco
19	344.5	18.8	521	21	AAA29152	Rice gamma-tocophe
20	344.5	18.8	521	21	AAA29162	Corn tassal-deride
21	316	17.3	311	24	ABL74098	Complete nucleotid
22	287	15.7	53500	21	AAC55842	S. lavenulae MltM
23	280.5	15.3	852	21	AAC55802	Physcomitrella pat
24	267.5	14.6	464	21	AAA29163	S. longisporoflavus
25	265	14.5	488	22	AAHA4245	Oxidoreducing aver
26	261	14.3	2122	18	AAE70152	A. halophilla betel
27	236	12.9	1410	15	AAO80524	Rice gamma tocophe
28	235	12.7	2814	21	AAZ50960	Maize C-24 sterol
29	228.5	12.5	592	21	AAA28153	Arabidopsis thalia
30	227	12.4	1383	21	AAZ92614	E. halochloridis bet
31	226	12.3	3219	21	AAZ50961	S. longisporoflavus
32	225	12.3	3219	21	AAZ50961	Corn SMT cDNA. Ze
33	225	12.3	6085	18	AAE70153	S. spinosa DNA fra
34	224.5	12.3	1497	20	AAV70839	S. spinosa DNA fra
35	220.5	12.0	828	22	AAE88318	DNA fragment of Sa
36	220.5	12.0	45624	22	AAE88315	Nicotiana tabacum
37	220.5	12.0	50000	22	AAE88312	Arabidopsis thalia
38	220.5	12.0	80161	20	AAZ21501	Arabidopsis thalia
39	218	11.9	12381	21	AAZ58381	Arabidopsis thalia
40	207.5	11.3	1041	24	AAZ23672	Arabidopsis thalia
41	204.5	11.2	1233	21	AAE48513	Arabidopsis thalia
42	203.5	11.1	1235	21	AAE39342	Arabidopsis thalia
43	203.5	11.1	1270	21	AAE48512	Arabidopsis thalia
44	203.5	11.1	1274	21	AAE39297	Arabidopsis thalia
45	203.5	11.1	1320	20	AAV70837	Yeast SMT gene. S

ALIGNMENTS

RESULT 1
ID AAA29164 standard; cDNA; 1189 BP.
AAA29164:
12-SEP-2000 (first entry)

XX Soybean gamma-tocopherol methyltransferase cDNA (clone sahlc.pK004.g2).
XX Vitamin E; alpha-tocopherol; biosynthesis; enzyme; inhibitor; herbicide;
XX gamma-tocopherol methyltransferase; ss.
XX Glycine max.
XX Key
XX Location/Qualifiers
XX 5..1057
XX CDS
XX /tag= a

```

FT XX /product= gamma-tocopherol_methyltransferase
PN XX WO200032757-A2.
PD XX 08-JUN-2000.
PF XX 02-DEC-1999; 99WO-US28588.
PR XX 03-DEC-1998; 98US-0110781.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX PI Cahoon RE, Coughlan SJ, Miao G, Rafalski JA:
XX
XX WPI: 2000-412309/35.
XX P-PDB: AAY96472.
XX
XX Polyunsaturated encoding a vitamin E biosynthetic enzyme polypeptide
XX such as gamma tocopherol methyltransferase, useful for synthesizing
XX vitamin E or as a target for design and discovery of herbicides
XX
XX Claim 3: Page 64-65; .82pp: English.
XX
XX AAA29151-69 encode vitamin E (alpha-tocopherol) biosynthetic enzymes or
XX their fragments. The cDNA clones were identified by BLAST searches based
XX on similarity to Synecocystis sp. and Arabidopsis thaliana vitamin E
XX biosynthetic enzymes. The enzymes are useful for synthesizing
XX plastoquinones or tocopherols (especially vitamin E). Vitamin E is
XX required for plant growth, therefore the enzymes may be used for the
XX discovery of new herbicides. The enzymes can be used in methods to
XX evaluate potential inhibitors, which may have use as herbicides.
XX Additionally, transgenic expression of, e.g. gamma-tocopherol
XX methyltransferase affords the ability to manipulate tocopherol levels as
XX desired for a particular application.
XX
XX Sequence 1189 BP; 333 A; 253 C; 299 G; 304 T; 0 other:
XX
XX
XX Alignment Scores:
XX Pred. No.: 2,34e-180 Length: 1189
XX Score: 1830.00 Matches: 350
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 21 Gaps: 0
XX
XX US-09-857-613a-28 (1-350) x AAA29164 (1-1189)
XX
XX 1 MetalathValAlaArgIleProThrIleSerCysIleHisIleHisThrPheArgSer 20
XX |||||||
XX 5 ATGCCACCGTGGTGGAGATCCCAACATCTCATCCATCCACATCCACCGTCCGTTCC 64
XX
XX 21 GluSerProAlaGlnPheAlaArgIleArgValGlyProAlaGlnSerTrpAlaProIleArg 40
XX |||||||
XX 65 CAATCCCTCCCTCCATTCGCCAGATCCGCGCGAGCCAGCTCTGCGCTCTTAATTCGG 124
XX
XX 41 AlaSerAlaAlaSerSerGluArgGlyIleValIleGluGlnIleProLysAsp 60
XX |||||||
XX 125 GCATCGCGAGGAGCTCCGAGAGAGGAGATGATATTGGACCAAGCCGGAAGAGGAT 184
XX
XX 61 AspLysLysLysLeuGlnLysGlyIleAlaGluPheTrpAspGluSerSerLysLeuTrp 80
XX |||||||
XX 185 GACAAAGAAAGAGCTGCACAAAGGATCCAGAGTTTACAGAGATCGTCTGGCTTATCG 244
XX
XX 81 GluAsnIleTrpGlyAspHisMetHisLysGlyPheTrpAspSerAspSerThrValSer 100
XX |||||||
XX 245 GAGAACATTTGGGGGACACATGCACATGCTTTTATGATCGCATCCACTCTTTCG 304
XX
XX 101 LeuSerAspHisArgAlaAlaGlnIleArgMetIleGlnGluSerLeuAlaPheAlaSer 120
XX |||||||
XX 305 CTTTCGATCAATCGCTGCTCAGATCCGATGATCCAGAGATCTCTTCGCTTGGCTCT 364
XX
XX 121 ValSerGluGlnArgSerLysTrpProLysSerIleValAspValGlyCysGlyIleGly 140
XX |||||||

```

```

Db 365 GTTTCGAGGAGCGTAGTAATGGCCCAAGATATAGTTGATGTGGGTGGCATAGCT 424
QY 141 GlycerSerArgTrpIleAlaAlaLysLysPheGlyAlaThrSerValGlyIleHisLeuSer 160
Db |||||||
Db 425 GGCAGCTCTGATACCTGGCCAGAGAAATTTGGACCAACCTGTAGGCATCTAGT 484
QY 161 ProValGlnAlaGlnArgAlaAsnAlaLeuAlaAlaAlaGlnGlyLeuAlaAspLysVal 180
Db |||||||
Db 485 CTTGTTCAAGCTCAAAAGAGCAATGCTGTGCTGCTGCTGCTCAAGATGGCTGATTAAGTT 544
QY 181 SerPheGlnValAlaAspAlaLeuGlnIleProPheSerAspGlyGlnPheAspLeuVal 200
Db |||||||
Db 545 TCTTTTCAAGCTTGTGAGCGCTCTACACCAACCATCTCTGTGACGGCAGTTTGATCTGGTG 604
QY 201 TrpSerMetGluSerGlyGlnHisMetProAspLysAlaLysPheValGlyLeuAla 220
Db |||||||
Db 605 TGTGTCATGAGAGATGGAGCATATGCTGACAAAGCTTAAGTTTGGAGAGATTAGCT 664
QY 221 ArgValAlaAlaProGlyAlaIleIleIleValThrTrpCysHisArgAspLeuGly 240
Db |||||||
Db 665 CGGTTAGAGGACGACGAGTGCATATATATATATATATATATATATATATATATATATATAT 724
QY 241 ProAspGluGlnSerLeuHisProTrpGluGlnAspLeuLysLysIleCysAspAla 260
Db |||||||
Db 725 CCTGACGACATCTTACATCATGAGGAGCAGATCTTAAAGAGATTGGCATGCA 784
QY 261 TyrTrpLeuProAlaTrpCysSerThrSerAspTrpValLysLeuGlnSerLeuSer 280
Db |||||||
Db 785 TATTACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 844
QY 281 LeuGlnAspIleLysSerGluAspTrpSerArgPheValAlaProPheTrpProAlaVal 300
Db |||||||
Db 845 CTTGACGACATCAAGTGAAGATGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 904
QY 301 IleArgSerAlaPheThrTrpLysGlyLeuSerSerLeuLeuSerSerGlyGlnLysThr 320
Db |||||||
Db 905 ATAGCTCAGCTTCACATGAGAGGCTATCTTCCATCTTCCATCTTCCATCTTCCATCTTCCATCTT 964
QY 321 IleLysGlyAlaLeuAlaMetProLeuMetIleGluGlyTrpLysAspLeuLysThr 340
Db |||||||
Db 965 ATAAAGAGAGCTTTGGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 1024
QY 341 PheAlaIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 360
Db |||||||
Db 1025 TTTGCCATCATTTACATGTGCAAAACCTGAA 1054
XX
XX RESULT 2
XX AAA29154
XX ID AAA29154 standard; cDNA: 1331 BP.
XX AC AAA29154;
XX
XX 12-SEP-2000 (first entry)
XX
XX Soybean gamma tocopherol methyltransferase contig cDNA.
XX
XX Vitamin E: alpha-tocopherol; biosynthesis; enzyme; inhibitor; herbicide;
XX gamma-tocopherol methyltransferase; ss.
XX
XX Glycine max.
XX
XX Key Location/Qualifiers
XX CDS 5..1055
XX FT /*tag= a
XX FT /product= gamma_tocopherol_methyltransferase
XX
XX WO200032757-A2.
XX
XX 08-JUN-2000.
XX
XX 02-DEC-1999; 99WO-US28588.
XX
XX 03-DEC-1998; 98US-0110781.
XX

```

XX (DUP0) DU PONT DE NEMOURS & CO E.I.
 PA Cahoone RE, Coughlan SJ, Miao G, Rafalski JA:
 PI WPI: 2000-412309/35.
 DR P-PSDB: AAY96462.
 XX
 PT Polynucleotide encoding a vitamin E biosynthetic enzyme polypeptide
 PT such as gamma tocopherol methyltransferase, useful for synthesizing
 PT vitamin E or as a target for design and discovery of herbicides
 PS
 PS Claim 3: Page 52: 82pp: English.
 XX
 CC AAA29151-69 encode vitamin E (alpha-tocopherol) biosynthetic enzymes or
 CC their fragments. The cDNA clones were identified by BLAST searches based
 CC on similarity to Synecocystis sp. and Arabidopsis thaliana vitamin E
 CC biosynthetic enzymes. The enzymes are useful for synthesizing
 CC plastoquinones or tocopherols (especially vitamin E). Vitamin E is
 CC required for plant growth, therefore the enzymes may be used for the
 CC discovery of new herbicides. The enzymes can be used in methods to
 CC evaluate potential inhibitors, which may have use as herbicides.
 CC Additionally, transgenic expression of, e.g. gamma-tocopherol
 CC methyltransferase affords the ability to manipulate tocopherol levels as
 CC desired for a particular application.
 XX
 SO Sequence 1331 BP: 373 A: 293 C: 301 G: 364 T: 0 other:
 Alignment Scores:
 Pred. No.: 1.66e-163 Length: 1331
 Score: 1668.50 Matches: 323
 Percent Similarity: 94.52% Conservative: 5
 Best Local Similarity: 93.08% Mismatches: 10
 Query Match: 91.17% Indels: 9
 DB: 21 Gaps: 1
 US-09-857-613A-28 (1-350) x AAA29154 (1-1331)
 QY 1 MetalAlThrValAlaArgLeuProThrIleSerCysIleHisIleHisThrPheArgSer 20
 DB 6 ATGGCCACCGGTGTGAGGATCCCAACAAATCTCATGCTCCACATCCACAGCTTCC 65
 QY 21 GlnSerProArgThrPheAlaArgIleArgValGlyProArgSerTrpAlaProIleArg 40
 DB 66 CAATCCCTCCGACCTTCCGCCAGAAATCCGGCTCGACCCAGCTCGGCTCTATTCCG 125
 QY 41 AlaSerAlaAlaSerSerGluArgGlyIleValIleValLeuGluGlnIleProLysLysAsp 60
 DB 126 GCATCGCGAGCGAGCTCGAGAGAGGCGAGATGATGATGAGCAGCAAGCCGAAAGAGAT 185
 QY 61 AsplyLysLysLeuGlnLysGlyIleAlaGluPheTrpAspLysSerGlyLeuTrp 80
 DB 186 GCAAGAAAGAACTCCAGAGGAAGCAAGATTTTACACAGAGCTGCTGCTTATG 245
 QY 81 GluAlnIleTrpGlyAspHisMetHisIleGlyPheTrpAspSerAspSerThrValSer 100
 DB 246 GAGAACATTTGGGGCGACCATGACCATGCTTTTATGACTCGGATTCCTGCTTTG 305
 QY 101 LeuSerAspHisArgAlaAlaGlnIleArgMetIleGlnIleSerLeuArgPheAlaSer 120
 DB 306 CTTTCGAGATCATGCTGCTCCTCAGATCCGATGATCCAAAGATCTTTCCTTGCCTCT 365
 QY 121 ValSerGluGluArgSerLysTrpProLysSerIleValAspAlaGlyCysGlyIleGly 140
 DB 366 GTTTCGAGAGAGCTAGTAAATGGCCCAAGAGATGATGATGCTTGGTGTGCAATAGCT 425
 QY 141 GlySerSerArgTrpLeuAlaLysLysPheGlyAlaThrSerValGlyIleThrLeuSer 160
 DB 426 GGCAGCTCTAGTACCTGGCCCAAGAAATTTGGGCAACCAAGCTTAGCATCCTGAGT 485
 QY 161 ProValGlnAlaGlnArgAlaAsnAlaLeuAlaAlaGlnIleLeuAlaAspLysVal 180
 DB 486 CCGTTCAGAGCTCAAGAGCAAAATGCTTGTGCTGCTCAAGAGATTGGCTGATTAAGGTT 545

QY 181 SerPheGlnValAlaAspAlaLeuGlnGlnProPheSerAspGlyGlnPheAspLeuVal 200
 DB 546 TCCCTTCAGCTGTGCGAGCCCTTACAGACACCATCTCTGAGAGCGCAGTTGATCTGGTG 605
 QY 201 TrpSerMetGluSerGlyGluHisMetProAspLysAlaLysPheValGlyLeuAla 220
 DB 606 TGGTCCATGAGAGTGGAGAGCATATGCTGCAAAAGTAAAGTTGTTGGACAGTTAGCT 665
 QY 221 ArgValAlaAlaProGlyAlaIleIleIleIleValIleThrCysHisArgAspLeuGly 240
 DB 666 CCGGTAGACAGCACAGGTGCCATTTAATAATAGTAACATGGTCCACAGGATCTTGGC 725
 QY 241 ProAspGluGlnSerLeuHisProTrpGluGlnAspLeuLysLysIleCysAspAla 260
 DB 726 CCGAGACACATCTCTTACATCATGAGCAAGATCTTTAAAGAAAGTTTGGCATGCA 785
 QY 261 TyrTrpLeuProAlaTrpCysSerThrSerAspTrpValLysLeuGlnSerLeuSer 280
 DB 786 TATTACCTCCCTGCGTGGCTCAACTTCGATTAATGTTAAGTTGCTCCAAATCCCTGTCA 845
 QY 281 LeuGlnAspIleLysSerLysAspTrpSerArgPheValAlaProPheTrpProAlaVal 300
 DB 846 CTTTCAGACATCAAGTCAAGAAATGCTCGCTTGTGCTCCATTTTGGCCAGCAGTG 905
 QY 301 IleArgSerAlaPheThrTrpLysGlyLeuSerSerLeuLeuSerSerGlyGlnLysTrp 320
 DB 906 ATTCGCTCAGCTTCCATCATGAGAAAGGCTTATCTTCTTACGATGAGTGAAGCTTGA 965
 QY 321 IleLysGlyAlaLeuAlaMetProLeuMetIleGlyIleTyrLysLysAspLeuIleLys 340
 DB 966 ATTATATATGCA-----TTTCAAAACAAACCCCCCA 998
 QY 341 PheAlaIleIleIleThrCysArg 347
 DB 999 TCTTCTATTGCAACTTGCAAG 1019
 RESULT 3
 AAA29165
 ID AAA29165 standard: cDNA: 1257 BP.
 XX
 AC AAA29165:
 XX
 DT 12-SEP-2000 (first entry)
 XX
 DE Soybean gamma-tocopherol methyltransferase cDNA.
 XX
 KW Vitamin E: alpha-tocopherol; biosynthesis; enzyme; inhibitor; herbicide;
 KW gamma-tocopherol methyltransferase; ss.
 XX
 OS Triticum aestivum.
 XX
 XX
 FH Key
 FT CDS Location/Qualifiers
 FT 18..1118
 FT /tag= a
 FT /transl_except= (pos:30..32, aa:xaa)
 FT /product= gamma-tocopherol_methyltransferase
 FT /note= "Xaa is not defined"
 PN
 FN
 XX
 PD 08-JUN-2000.
 XX
 PF 02-DEC-1999: 99WO-US28588.
 XX
 PR 03-DEC-1998: 98US-0110781.
 XX
 PA (DUP0) DU PONT DE NEMOURS & CO E. I.
 PI Cahoone RE, Coughlan SJ, Miao G, Rafalski JA:
 XX WPI: 2000-412309/35.

DR P-PSDB: AAY96473.
 XX Polynucleotide encoding a vitamin E biosynthetic enzyme polypeptide
 PT such as gamma tocopherol methyltransferase, useful for synthesizing
 PT vitamin E or as a target for design and discovery of herbicides
 XX
 PS Claim 3, Page 66; 82pp; English.
 XX
 CC AAA29151-69 encode vitamin E (alpha-tocopherol) biosynthetic enzymes or
 CC their fragments. The cDNA clones were identified by BLAST searches based
 CC on similarity to *Synechocystis* sp. and *Arabidopsis thaliana* vitamin E
 CC biosynthetic enzymes. The enzymes are useful for synthesizing
 CC plastoquinones or tocopherols (especially vitamin E). Vitamin E is
 CC required for plant growth, therefore the enzymes may be used for the
 CC discovery of new herbicides. The enzymes can be used in methods to
 CC evaluate potential inhibitors, which may have use as herbicides.
 CC Additionally, transgenic expression of, e.g., gamma-tocopherol
 CC methyltransferase affords the ability to manipulate tocopherol levels as
 CC desired for a particular application.
 CC
 XX
 SQ Sequence 1257 BP; 299 A; 370 C; 343 G; 243 T; 2 other:
 Alignment Scores:
 Pred. No.: 4.76e-112 Length: 1257
 Score: 1173.00 Matches: 224
 Percent Similarity: 80.82% Conservative: 33
 Best Local Similarity: 70.44% Mismatches: 61
 Query Match: 64.10% Indels: 0
 DB: 21 Gaps: 0
 US-09-857-613A-28 (1-350) x AAA29165 (1-1257)
 QY 33 ProAAGSerTTPAlaProLeuArgAlaSerAlaSerSerGluArgGlyLeuVal 52
 DB 117 CCGCGGCTGCTCCCTCCCTCCCGGAGACAGCTCCGCGCCCGCATGCGCTGC 176
 QY 53 LeuGluGlnLysProLysLysAspLysLysLysLysLysLysLysLysLysLys 72
 DB 177 GACGACCGC 236
 QY 73 TyrAspGluSerSerGlyLeuTrpLysLysLysLysLysLysLysLysLysLys 92
 DB 237 TACGACGAGTCTCCGCGCTGCGGAGACATCTGGGCGGACATGCACACCGCTTC 296
 QY 93 TyrAspSerAspSerThrValSerLeuSerAspHisArgAlaGlnLeuArgMetIle 112
 DB 297 TACGACTCCGCGGAGCGCGCGCTCCATGTCGACACCGCGCGCGCGCATGCCATGATC 356
 QY 113 GlnGluSerLeuArgPheAlaSerValSerGluLysArgSerLysTrpProLysSerIle 132
 DB 357 GAGGAGCGCGCTCCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 416
 QY 133 ValAspValGlyCysGlyIleGlyLysSerSerArgTyrLeuAlaLysLysPheGlyAla 152
 DB 417 GTTATGATGTTGATGCGGATCGGTGTGATCTCAAGATACCTGGCAAAATATGAGCA 476
 QY 153 ThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaAsnAlaLeuAlaIle 172
 DB 477 CAATGCTCTGGATCATCATTTAGAGCCAGAGTGAAGAGAGAAATGCCCCGCGGCA 536
 QY 173 AlaGlnGlyLeuAlaLysValSerPheGlnValAlaAspAlaLeuGlnGlnIleProPhe 192
 DB 537 GCGGAGGCGGTTCGACAGAGCGCTTCTCAAGTGTGATGCTGTGGGCAACCATTT 596
 QY 193 SerAspGlyLysPheAspLeuValTrpSerMetGluSerGlyLysIleMetProAspLys 212
 DB 597 CCTGATGGGCGAGTTGATCTTGTCTGTATGAGAGAGTGTGAGCAATGCGGACAA 656
 QY 213 AlaLysPheValGlyGluLeuAlaArgValAlaAlaProGlyAlaIleIleIleVal 232
 DB 657 CAGAACTTTGTAAGCGATGCGACCGCTCCGAGCTCCAGGACCAATCATCATCATC 716
 QY 233 ThrTrpCysHisArgAspLeuGlyProAspGluGlnSerLeuHisProTrpGluGlnAsp 252

DB 717 ACCTGATGCCATAGAGAACTCCGCGCGGAGGAGACTCATGAAACCTGACGAGCTGAT 776
 QY 253 LeuLeuLysLysLysLysAspAlaTyrTyrLeuProAlaTrpCysSerThrSerAspTyr 272
 DB 777 CTTTGAAGAAAGATTGTGATGATATTCCTCCGAGATGTGCTCGCCCTCGGATTAAT 836
 QY 273 ValLysLeuLeuGlnSerLeuSerLeuGlnAspIleLysSerGluLysTrpSerArgPhe 292
 DB 837 GTCAAGATTGCCGAGCATTTCTCTTGGAGATATCAAAACGCCGAGCTGCAGAAAC 896
 QY 293 ValAlaProPheTrpProAlaValIleArgSerAlaPheThrTrpLysGlyLeuSerSer 312
 DB 897 GTGGCCCGCTTCGCGCTGCTGCTGATCAATCAGACAGTCAAGAGAAAGCCCTCACTTCT 956
 QY 313 LeuLeuSerSerGlyLysTrpTrpLysGlyAlaLeuAlaMetProLeuMetIleGlu 332
 DB 957 CTACTAAGAGAGTGAAGAGAGAGAAAGGAGGACCTGTGTATGCTTCATGATCCAA 1016
 QY 333 GlyTyrLysLysAspLeuLysPheAlaIleIleThrCysArgLysProGlu 350
 DB 1017 GCGTACAGAAAGCGCTCATTAAGTTACAGCATCATCCTGCGGCAAAACCCCA 1070
 RESULT 4
 AAA29161
 ID AAA29161 standard; cDNA; 1102 BP.
 XX
 AC AAA29161:
 XX
 DT 12-SEP-2000 (first entry)
 XX
 DE Corn gamma-tocopherol methyltransferase cDNA.
 XX
 KW Vitamin E; alpha-tocopherol; biosynthesis; enzyme; inhibitor; herbicide;
 KW gamma-tocopherol methyltransferase; ss.
 XX
 OS Zea mays.
 XX
 FT Key Location/Qualifiers
 FT CDS 1..1059
 FT /tag a
 FT /product- gamma-tocopherol_methyltransferase
 FT /transl_except- (pos:454..456, aa:Xaa)
 FT /note- "Xaa not defined"
 PN MO200032757-A2.
 XX
 PD 08-JUN-2000.
 XX
 PF 02-DEC-1999; 99WO-US28588.
 XX
 PR 03-DEC-1998; 98US-0110781.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Cahoon RE, Coughlan SJ, Miao G, Rafalski JA;
 DR WPI: 2000-412309/35.
 DR P-PSDB: AAY96469.
 XX
 PT polynucleotide encoding a vitamin E biosynthetic enzyme polypeptide
 PT such as gamma tocopherol methyltransferase, useful for synthesizing
 PT vitamin E or as a target for design and discovery of herbicides
 XX
 PS Claim 3, Page 60; 82pp; English.
 XX
 CC AAA29151-69 encode vitamin E (alpha-tocopherol) biosynthetic enzymes or
 CC their fragments. The cDNA clones were identified by BLAST searches based
 CC on similarity to *Synechocystis* sp. and *Arabidopsis thaliana* vitamin E
 CC biosynthetic enzymes. The enzymes are useful for synthesizing
 CC plastoquinones or tocopherols (especially vitamin E). Vitamin E is
 CC required for plant growth, therefore the enzymes may be used for the
 CC discovery of new herbicides. The enzymes can be used in methods to

CC evaluate potential inhibitors, which may have use as herbicides.
CC Additionally, transgene expression of, e.g. gamma-tocopherol
CC methyltransferase affords the ability to manipulate tocopherol levels as
CC desired for a particular application.

XX Sequence 1102 bp, 231 A; 330 C; 328 G; 207 T; 6 other:

Alignment Scores:

Score:	5.5e-111	Length:	1102
Percent Similarity:	1162.00	Matches:	225
Best Local Similarity:	75.88%	Conservative:	33
Query Match:	66.18%	Mismatches:	72
	63.50%	Indels:	10
		Gaps:	2

US-09-857-613A-28 (1-350) x AAA29161 (1-1102)

```
QY 20 SerGlnSerProArgThrPheAlaArgIleArgValGlyProArgSerTrpAlaProIle 39
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 28 TCCGAGTCTCCAGAGCCCTCCAGCTCCGCGCCGCGAGCTACCGCCGCTCCG 87
QY 40 ArgAlaSerAlaAlaSerSerGlnArgGlyIleValLeuGlnGlnLysProLys 59
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 88 CACGTCGCCGCCACTCCCGCGCTCCGAGCGCGCTGCTAGCTGCGCTCGAGGCC 147
QY 60 AspAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 71
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 148 TCGTCGAGCGGCTCAGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 207
QY 72 PheTrpAspGlnSerSerGlyLeuTrpGlnAsnIleTrpGlyAspHisMetHisGly 91
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 208 CTGTAACGACGAGTCTCGGGGCTGGGAGAACATCTGGCGCGCACATGACACAGCC 267
QY 92 PheTrpAspSerAspSerThrValSerLeuSerAspHisArgAlaAlaGlnIleArgMet 111
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 268 TTCTACGACTCGAGCGAGCGCGCTCCATGCGCATCGCGCGCGCGCGCGCATG 327
QY 112 IleGlnGlnSerLeuArgPheAlaSerVal-----SerGlnGlnArgSerLysTrpPro 129
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 328 ATCGAGGAGCGCGCTCGCTCCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCG 387
QY 130 LysSerIleValAspValGlyCysGlyIleGlyGlySerSerArgTrpLeuAlaLysLys 149
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 388 AAACAAATAGTCATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTC 447
QY 150 PheGlyAlaThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaAsnAla 169
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 448 TACGAGACGAGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATG 507
QY 170 LeuAlaAlaAlaGlnGlyLeuAlaAspLysValSerPheGlnValAlaAspAlaLeuGln 189
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 508 CTCGCTGAGCGCGAGGGTTGTCGATGTCGATGTCGATGTCGATGTCGATGTCG 567
QY 190 GlnProPheSerAspGlyGlnPheAspLeuValTrpSerMetGlnSerGlyLysHisMet 209
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 568 CAACCGTTTCCGAGCGCGAGTTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCG 627
QY 210 ProAspLysAlaLysPheValGlyIleLeuAlaArgValAlaAlaProGlyAlaIleIle 229
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 628 CCGGCAAGAGAAAGTTGTTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 687
QY 230 IleIleValThrTrpCysHisArgAspLeuGlyProAspGlnGlnSerLeuHisProTrp 249
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 688 ATCATCGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGAT 747
QY 250 GlnGlnAspLeuLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLys 269
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 748 GAACGACGCTCTCGAGAGATATGACGCGGTACTACTCCGAGCTGGTCTACCT 807
QY 270 SerAspTrpValLysLeuLeuGlnSerLeuSerLeuGlnAspIleLysSerLysTrp 289
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 808 TCAGACTATGTGAACTTGCACAAATGCTCTCTCGAGGATATCAAGACAGCTGACTG 867
```

```
QY 290 SerArgPheValAlaProPheTrpProAlaValIleArgSerAlaPheThrTrpLysGly 309
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 868 TCGAGAACGTGCGCCCGCTTTTGCCCGCGCGATGAATCAACCGCTACATGAGAGGCC 927
QY 310 LeuSerSerLeuLeuSerSerGlyGlnLysThrIleLysGlyAlaLeuAlaMetProLeu 329
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 928 TTCACCTCTCTGTGACGACCGCATGATGAGAGCATGAGAGCGCATGTGATGCGCTA 987
QY 330 MetIleGlnGlyTrpLysLysAspLeuIleLysPheAlaIleIleThrCysArgLysPro 349
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 988 ATGATTCAGGCGCTACAGAGAGGCGCTCATCAATTCACCATCATCATCTGTCAGACCT 1047
RESULT 5
AAC50535
ID AAC50535 strand: DNA; 1293 bp.
XX AAC50535:
AC AAC50535:
XX 18-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 65205.
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 65205.
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX BP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX PF
XX 25-FEB-1999; 99US-0121825.
XX PR 09-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0130891.
XX PR 28-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
```

PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 4,15e-110
Score: 1154.50
Percent Similarity: 75.43%

Length: 1293
Matches: 231
Conservative: 30


```

180 GAC-----CGCTAAGA 191
OY 67 LysGlyLeuIleAclunpTherYrAspGluSerSerGlyLeuTrpGluAsnIleTrpGlyAsp 86
DB 192 AAGAGATGACGGAGTTCTCAAAATGAAACTTCGGGTTTGGAAGAGATTTGGGAGAT 251
OY 87 HisMetHisIleGlyPheTherYrAspSerSerTherValSerLeuSerAsp-----His 104
DB 252 CAAATCATCATGCGCTTTATATGACCGCTGATTCCTGTTCAACTTCTGATTTCTGTGTCAC 311
OY 105 ArgAlaAlaGlnIleArgMetIleGlnGluSerLeuArgPheAlaSerValSer---Glu 123
DB 312 AAGAGAGCTCAGATCCGATGATGAAGAGTCTCCGTTCCCGCGTGTACTGATGA 371
OY 124 GluArgSerLysTrpProLysSerIleValAspValGlyCysGlyIleGlyGlySerSer 143
DB 372 GAGGAGGAGAAAGATAAAGAAAGATGATGTTGGGATGGGATGAGAGAACTCA 431
OY 144 ArgTrpLeuAlaLysLysPheGlyAlaThrSerValGlyLeuThrLeuSerProValGln 163
DB 432 AGATATCTTGCCCTTAATTTGAGAGCTGAATGATGCGCTTACTCTCAGCCCTGTTCAG 491
OY 164 AlaGlnArgAlaAsnAlaLeuAlaAlaGlnGlyLeuAlaLysPheValSerPheGln 183
DB 492 GCCAAGAGAGCCAAATGATCTCGCGCTGCTCAATCACTCTCATAGAGCTTCTCCAA 551
OY 184 ValAlaAspAlaLeuGlnGlnTrpPheSerAspGlyGlnPheAspLeuValTrpSerMet 203
DB 552 GTTGGCGGATGCGTGGATGATGATTCGAAGATGAAGAAATTCATCTAGTGTGATG 611
OY 204 GluSerGlyLysIleMetProAspLysAlaLysPheValGlyLeuLeuAlaArgValAla 223
DB 612 GAGAGTGGAGATATGCTGCTGACAGGCCAATTTGTAAGAGTTCGACTGTGCGG 671
OY 224 AlaProGlyAlaIleIleIleValIleThrTrpCysHisArgAspLeuGlyProAspGlu 243
DB 672 GCTCCAGAGATGATGATTAATGATGACATGCTGCCATAGAATCTATCTCGCGGAG 731
OY 244 GlnSerLeuHisProTrpGluGlnAspLeuLeuLysIleCysAspAlaTrpTrpLeu 263
DB 732 GAAGCTTTCAGCCGCTGGAGCAAAACATCTTGACAAAATCTGTAAAGCGTTCTATCTC 791
OY 264 ProAlaTrpCysSerThrSerAspTrpValLysLeuLeuGlnSerLeuSerLeuGlnAsp 283
DB 792 CCGGCTTGGGTCCACACGATGATATGCACTTGTCTCAATCCATCTCTCCAGGAT 851
OY 284 IleLysSerGluAspTrpSerArgPheValAlaProPheTrpProAlaValIleArgSer 303
DB 852 ATTAAGTGTGCGGATGCTGAGAGAACGTAGCTCTTCTGCGCTGCGGTATACGAGCT 911
OY 304 AlaPheThrTrpLysGlyLeuSerSerLeuLeuSerSerGlyLysIleThrIleLysGly 323
DB 912 GCAATTACATGGAAGGCGCTGTGCTCTGCTGCTGATGAGTGAAGATTAAGA 971
OY 324 AlaLeuAlaMetProLeuMetIleGlnGlyTrpLysLysAspLeuIleLysPheAlaIle 343
DB 972 GCATTACAAATGCCATTGATGATGAAGTTACAAAGAAAGGTCTCATTAAGTTGGTATC 1031
OY 344 IleThrCysArgLysPro 349
DB 1032 ATCACTTCCAGAACCA 1049

```

RESULT 7
AAL46039
ID AAL46039 standard: cDNA: 1350 BP.

AC AAL46039;
XX 11-JUL-2002 (first entry)
DT A thaliana gamma-tocopherol methyltransferase cDNA.
DE XX

```

KW Vitamin E; homogenitase; HG: homogenitase-1,2-dioxygenase; HGD:
KW maleylacetate-1-isomerase; MAAI: fumarylacetate-hydrolase; FAH:
KW transgenic; nutrition; antioxidant; cardiovascular disease; cancer;
KW immune system; generalised age-related degeneration; animal feed;
KW meat quality; cosmetics; growth regulator; herbicide; cardiant;
KW cytosolic; immunostimulant; enzyme; gene; ss.
OS Arabidopsis thaliana.
XX
XX
FH key Location/Qualifiers
FT CDS 63..1109
FT /tag= a
FT /product= "gamma-tocopherol methyltransferase"
PN MO200231173-A2.
XX
XX
PD 18-APR-2002.
XX
XX
PF 18-SEP-2001; 2001WO-BP10779.
XX
XX
PR 19-SEP-2000; 2000DE-1046462.
XX
XX
PA (SUNG-) SUNGENE GMBH & CO KGAA.
XX
XX
PI Geiger M, Ebneith M, Kunze I.
XX
XX
DR MPI: 2002-362682/39.
XX
XX
DR P-PSDB: AAO17426.
XX
XX
PT Increasing Vitamin E production, useful in human or animal nutrition,
PT comprising reducing activity of enzymes that metabolize homogenitase
XX
XX
PS Disclosure: Page 83-85; 98pp; German.
XX
XX
CC The present invention relates to a method of producing Vitamin E,
CC involving modulating the synthesis of vitamin E by reducing decomposition
CC of homogenitase (HG) by reducing the activity of
CC homogenitase-1,2-dioxygenase (HGD), maleylacetate-1-isomerase (MAAI)
CC and/or fumarylacetate-hydrolase (FAH). The method is used to
CC produce transgenic organisms that are useful in animal and human
CC nutrition and for the isolation of vitamin E. Vitamin E is a fat-soluble
CC antioxidant with a protective effect against cardiovascular disease and
CC cancer, it also stimulates the immune system and may prevent generalised
CC age-related degeneration. When used in animal feeds, it improves quality
CC and storability of meat and can also be used in cosmetics. Also
CC antibodies raised against the 3 specified enzymes are useful in screening
CC for specific inhibitors, potentially useful as growth regulators, e.g.
CC herbicides. The present sequence is the A thaliana gamma-tocopherol
CC methyltransferase cDNA.
XX
XX
SQ Sequence 1350 BP; 374 A; 278 C; 314 G; 384 T; 0 other;

```

Alignment Scores: 9,04e-110 Length: 1350

Pred. No.: 1151.50 Matches: 230

Score: 75.43% Conservative: 65

Best Local Similarity: 66.47% Mismatches: 21

Query Match: 62.92% Indels: 21

DB: 24 Gaps: 4

US-09-857-613A-28 (1-350) x AAL46039 (1-1350)

OY 8 ProThiIleSerCysIleHisIleHisThrPheArgSerGlnSerProArgThrPheAla 27

DB 115 CCAACTCTTCTTGGCTCAAGATATCGCTTCTTTCGCTTCAATCCCTCTCTCC 173

OY 28 ArgIleArgVal---GlyProArgSerTrpAlaProIleArgAlaSerValSer 46

DB 174 TCAGTCTCTATACGACGACGCGTGAAGCGTGTGCGCTGCTCTCTCATTCACCT 233

OY 47 GluArgGlyGluIleValLeuGlnGlnLysProLysLysAspAspLysLysLysLeuGln 66

DB 234 GAG-----GCCTAAGA 245

QY	67	LYSGYLIALAIGLPHETHYRSPGLUSERSEGLYLEUTRPLUSNIETRPSGLYAS	86
Db	246	AAAGCAATACCGGAGTTCTTACATGAAACTTCGGGTTTGTGGAGAAGATTGGGAGAT	305
QY	87	HISMETHISHSLGYPHETHYRSPSERASPETHVALSERDEUSERASP-----HIS	104
Db	306	CATATGCATCATGCGCTTTTATATGACCCTGATTTCTTGTTCAACTTCTGTATCTGGTCAAC	365
QY	105	ARGALIALAGINILLEARGHETLLEGINGLUSERLEUARGPHEALASERVALSER---GLU	123
Db	366	AAGGAAGCTCAATCCGTAATGATTGAAGAGTCCTCCGTTGCGCCGCTGTACTGATTGAAGAA	425
QY	124	GUATGSELYSTRIPROLYSERLILEVALASVALGILCYSGLYLLEGLYGLYSERSET	143
Db	426	GAGGAGGAGGAAATAAGTAAAGAAAGTAGTGAGTGTGGTGGTGGATTGGAGGAACTCA	485
QY	144	ARGTYRLEUALALYSLYSPHEGLYALARTHRSERVALGILYLETHRLEUSERPROVALGIN	163
Db	486	AGATATCTTGCCCTCAATAATTTTGGACCTGTAATCGATGGCATTACTCTCAACCCCTGTTCAG	545
QY	164	ALAGINATGALIAASALALAEUALALALAGINGLYLEUALALASPLYVALSERPHEGIN	183
Db	546	GCCAAAGAGGCCAATGATATCTCCGGCTGCTCATCTCTCATATAGGCTTCCTTCCAA	605
QY	184	VALALASPALALAEUCLINGINIPROPHESERASPLYGLINPHEASPLEUVALTRPSERMET	203
Db	606	GTTGGGAGATGCGTTGGATGATGACCATTCGAAGATGGAATAATTCATCTTAGTGIGTGGATG	665
QY	204	GLUSERGLYGLINHISMETPROASPLYSALALYSPHEVALGLYGLULEUALARGVALALAL	223
Db	666	GAGAGTGTGGACATATAGCTCTACCAAGCCCAAGTTGTAAAGACAGTTGACTGTGGCGG	725
QY	224	ALAPROGLYALALILELLELLELLEVALTRTPTRPCYSHISARGASPLEUGLYPROASPLGI	243
Db	726	GCTCCAGGAGTAGCATATATATAGTACATGCTGCCATGAGAAATCTATATCTGGCGGGAG	785
QY	244	GLINSELEUHSIPROTTRPCLUGLINASPLEULEULYSLYSLIECYASAPALATRYTYRLEU	263
Db	786	GAAAGCTTGGACACCCGNGGAGCAAAAACATCTTGGACAAAATCTGTAAAGCGTTCTATC	845
QY	264	PROALATRPCYSERETHRSERSPRYVALYLVSLEULEGINSEURLEUSERLEUGINASP	283
Db	846	CCGCGTGTGGTCCACCATCATATTATATATGCAACTTCCTTCATATCCCATCTCTCCCGAGAT	905
QY	284	ILELYSSEGLIUSPTRPSERARPRPHEVALALAPROPHETRPROALVALILEASGSR	303
Db	906	ATTAACTGTGGCGATTGGTGCACAGAACGTAGCTCTTTTGTGGCTTCGGTTATATCCGACT	965
QY	304	ALAPHERTRPYLYSGLYLEUSERSERLEUUSERSERSELYGLINYSTTHRILETYSGLY	323
Db	966	GCATTAAACATGGAAGAGCCCTCTGTGCTCTGCTTCGTAAGTGATGAAGAAAGTATTAAGCA	1025
QY	324	ALALEUALAMEPROLEUMETLLEGLUGLYTYRILYSLYASPLEULEULEYSPHEALALLE	343
Db	1026	GCATTGACAATGCCATTGATGATTGAAGGTTACAAAGAAAGTGCTCATTAAGTTGGCTATC	1085
QY	344	ILETHRYSARGLYSPRO 349	
Db	1086	ATCACTTGCCAGAACCA 1103	
RESULT	8		
ID	AAAX17789		
XX	AAAX17789		
AC	AAAX17789;		
DT	21-MAY-1999	(first entry)	
XX	Arabidopsis gamma-tocopherol methyltransferase encoding gene.		
DE	Gamma-tocopherol methyltransferase: gamma-TMT; SLR0089; 165H577; meat;		
XX	transgenic plant; alpha-tocopherol; gamma-tocopherol; forage plant; ds		
KW			

XX	Arabidopsis sp.
OS	
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	207..1253
FT	/tag= "a"
FT	/gene= "16SH5T7"
FT	/product= "gamma-TMT"
FN	
PN	W09904622-A1.
XX	
PD	04-FEB-1999.
XX	
PE	22-JUL-1998; 98MO-US15137.
XX	
PR	17-JUL-1998; 98US-0053819.
PR	25-JUL-1997; 97US-0053819.
PR	26-JAN-1998; 98US-0072497.
PA	(UYNE-) UNIV NEVADA.
XX	
PI	Dellapenna D, Shintani DK;
DR	
DR	WPI: 1999-142458/12.
P-PSDB:	AAM95017.
PT	Newly isolated DNA fragment comprising a ^c-tocopherol (vitamin E)
PT	methyltransferase coding sequence - useful for producing
PT	^c-tocopherol, and transgenic plants, seeds and oils with an altered
PT	tocopherol profile
XX	
PS	Claim 2; Page 36-38; 46pp; English.
XX	
CC	The invention provides DNA sequences encoding gamma-tocopherol
CC	methyltransferases (gamma-TMT). The gamma-TMT encoding genes SLR0089
CC	and 16SH5T7 are isolated from Synchocystis and Arabidopsis species
CC	respectively. The DNA fragments are useful for producing transgenic
CC	plants containing a higher alpha-tocopherol:gamma-tocopherol ratio. They
CC	are also useful for producing alpha-tocopherol and plants with increased
CC	gamma-tocopherol, which may be useful in certain industries such as the
CC	meat industry e.g. for developing forage plants to feed animals. The
CC	production of transgenic plants (and seeds) with a higher alpha-
CC	tocopherol:gamma-tocopherol ratio, using the gamma TMTs will increase
CC	the level of alpha-tocopherol in the human diet, in addition to enhancing
CC	the stability and shelf life of plants and plant products. Increased
CC	levels of alpha-tocopherol will also increase meat quality and extend
CC	shelf life of post-processed meat products. Plants with a higher alpha-
CC	tocopherol:gamma-tocopherol ratio may also produce advantageous
CC	phenotypes. The present sequence represents the 16SH5T7 gene encoding the
CC	Arabidopsis gamma-TMT.
XX	
SQ	Sequence 1790 BP; 474 A; 404 C; 436 G; 476 T; 0 other;
Alignment Scores:	
Pred. No.:	1,35e-109 Length: 1790
Score:	1151.50 Matches: 230
Percent Similarity:	75.43% Conservative: 31
Best Local Similarity:	66.47% Mismatches: 65
Query Match:	62.92% Indels: 21
DB:	20 Gaps: 4
US-09-857-613A-28 (1-350) x AAX17789 (1-1790)	
OY	ProthrIleSerCysIleHisIleThrPheArgSerGlnSerProArgThrPheAla 27
Db	:::
259	CAACCTCTTTTCGGCGCTAAAGTCAATGCCTCTCTTT-CGGTCTCCACTCTCTCTCC 317
OY	ArgIleArgVal---GlyProArgSerTrpAlaProIleArgAlaSerAlaIleSer 46
Db	::: :::: :::: ::
318	TGAAGTCATATGACGACAACGCCGTGGAAACGTCGTGGCGCGCTGCCTACATCCACT 377
OY	GluArgGluIleValLeuCluGlnLysProLysLysAspAspLysLysLysLeuGln 66
Y	

Db 378 GAG-----GCCCTAGA 389
Qy 67 LysGlyLLeaLagLuehPyrAspGluSerSerGlyLeuTrpGluAsnIleTrpGlyAsp 86
Db 390 AAAGAAATACCGGATCTCAATGAAGAACTCGGTTTGCGAGAGATTGGGAGAT 449
Qy 87 HisMetHisIscGlyPheTrpAspSerSerThraValSerLeuSerAsp-----His 104
Db 450 CATATGATCATGCGCTTTATGACCTGATTTCTCTGTTCACTTCTTGATTCGGTCAC 509
Qy 105 ArgAlaIacInIleArgMetIleGluIleSerLeuArgPheAlaSerValSer---Glu 123
Db 510 AAGAACTAGATCCCTATGATGAGAGAGCTCTCCGTTCCGCGGTACTGATGAA 569
Qy 124 GluArgSerLysTrpProLysSerIleValAspValGlyCysGlyIleGlySerSer 143
Db 570 GAGGAGAGAAAGAAAGTAAAGTAAGTAGATGTGGCTGTGGATGAGGAGAGCTCA 629
Qy 144 ArgTrpLeuAlaLysLysPheGlyAlaThrSerValGlyIleThrLeuSerProValGlu 163
Db 630 AGATATCTTGCTTAATTTGAGCTGAATGCATTGGCATTACTCTACGCCCTTTCAG 689
Qy 164 AlaGlnArgAlaAsnAlaLeuAlaAlaIacInGlyLeuAlaAspLysValSerPheGlu 183
Db 690 GCCAAGAGAGCCATGATCTCGCGCTCTCAATCATCTCTCAAGAGCTCTCCAA 749
Qy 184 ValAlaAspAlaLeuGlnGlnProPheSerAspGlyGlnPheAspLeuValTrpSerMet 203
Db 750 GTTCGCGATGCTGGATCGATCCATTCAGAGTAAATTCGATCTGATGTGGTAGAG 809
Qy 204 GluSerGlyGluHisMetProAspLysAlaLysPheValGlyLeuAlaArgValAla 223
Db 810 GAGAGTGTGAGCATATCTGACAGAGCCAACTTTGTAAGAGAGTGGTACTGGCGCG 869
Qy 224 AlaProGlyAlaIleIleIleValIleThrCysHisArgAspLeuGlyProAspGlu 243
Db 870 GCTCAGAGAGCATGATATATAGTACATGCTGCCATAGAAATCTATCTCGCGGAG 929
Qy 244 GluSerLeuHisProTrpGluGlnAspLeuLysLysIleCysAspAlaTrpTrpLeu 263
Db 930 GAAGCTTTCACCGCTGGAGCAAAACATCTTGACAAATCTGTAAAGCGTTCTATCTC 989
Qy 264 ProAlaTrpCysSerThrSerAspTrpValLysLeuLeuGlnSerLeuSerLeuGlnAsp 283
Db 990 CCGGCTGTGCTCCACCATGATATATGTCACACTTCTCAATCCATCTCTCCAGAT 1049
Qy 284 IleLysSerGluAspTrpSerArgPheValAlaProPheTrpProAlaValIleArgSer 303
Db 1050 ATTAACTGCGCGATTTGTCAGAGACGTAAGCTCTTTCGGCTCGGTTATACGAGCT 1109
Qy 304 AlaPheThrTrpLysGlyLeuSerSerLeuLeuSerSerGlyGlnLysThrIleLysGly 323
Db 1110 GCATTAACTGAGAGGCGCTGTCTCTGCTCGTAGTGTGAAAGCTATTAAAGGA 1169
Qy 324 AlaLeuAlaMetProLeuMetIleGluGlyTrpLysLysAspLeuIleLysPheAlaIle 343
Db 1170 GCATTACATGCGCATGATGTAAGGTAAACAGAAAGCTGATTAAGTTTGGTATC 1229
Qy 344 IleThrCysArgLysPro 349
Db 1230 ATCACTTGCAGAAAGCA 1247

RESULT 9
AAC34022
ID AAC34022 standard; DNA: 1296 BP.
XX AAC34022:
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 5167.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW

KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EPI033405-A2.
PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139464.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.

```

PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142390.
PR 06-JUL-1999; 99US-0142505.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145219.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 28-JUL-1999; 99US-0145918.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.

```

```

PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156599.
PR 28-SEP-1999; 99US-0156599.
PR 29-SEP-1999; 99US-0156599.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 28-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Alignment Scores:

```

Pred. No.: 2.22e-109
Score: 1147.50
Percent Similarity: 75.438
Best Local Similarity: 66.188
Query Match: 62.708
DB: 21
Gaps: 4

```

US-09-857-613A-28 (1-350) x AAC34022 (1-1296)

```

QY 8 ProthrlleSerCysIleHisIleHsrhPheArgSerGlnSerProArgThrPheAla 27
DB 120 CCACCTCTTCTTCGCGTCAAGTCATCGCTTCTT CGGTCTCCACCTTCCTCC 178
QY 28 ArgIleArgVal--GlyProArgSerTrpAlaPoiIleArgAlaSerAlaLaserSer 46
DB 179 TCAGTCTATGACGACACGCGTGAACGTCGCGCGCGCTGCTACATCCACT 238
QY 47 GluArgGlyIuIleValLeuGlnGlnProIlySlyAspAspLysLysLeuGln 66
DB 239 GAG-----GCCCTAACA 250
QY 67 LysGlyIleAlaGluPheTyraSpGlnSerSerGlyLeuTrpGluAnIleTrpGlyAsp 86
DB 251 AAAGGAATACGCGAGTCTTCAATGAACATTCGCGTTGTGGGAAGATTTGGGAGAT 310
QY 87 HisMetHisHisGlyPheTyraSerAspSerThrValSerLeuSerAsp-----His 104

```



```

OY 119 AlaSerValSerGluGluValGlySerLysTrpProLysSerIleValAspValGlyCysGly 138
DB 559 GCTGGCGCTCTCTACCAAGATTGTAACCGAAGACGATCGTCGATGGGTTGGG 618
OY 139 ILGGLYSerSerArgTrpLeuAlaLeuLysPheGlyAlaThrSerValGlyIleThr 158
DB 619 ATAGGGGGAAGCTTCACGTTACTGGCCGGAATTCACGCGCAAGGTGATGCCATCAGC 678
OY 159 LeuSerProValGlnAlaGlnAlaAsnAlaLeuAlaLeuAlaGlnGlyLeuAlaAsp 178
DB 679 CTCAGGCCAGTGCAGGTTCAAGACGCCGTAAGACCTTACTGCAAGCAAGCTTATCTGAC 728
OY 179 LysValSerPheGlnValAlaAspAlaLeuGlnGlnProPheSerAspGlyGlnPheAsp 198
DB 739 CTCGTCATTTCCAGTACGGAATGCGCTCAAGCACCCCTTACGATGGTCTGTTGAT 798
OY 199 LeuValTrpSerMetGluSerGlyGluHisMetProAspLysAlaLysPheValGlyGlu 218
DB 799 CTCGCTGGTCCATGAGAGCGCGGACATGCCAGCAAGAAAGTTTGGGCGCAG 858
OY 219 LeuAlaArgValAlaAlaProGlyAlaIleIleIleIleValThrTrpCysHisArgAsp 238
DB 859 CTTCGACGAGTACAGCTCCGCGCGCTGCATTTATCTGTCAGCTGCGCCACCGTGTAT 918
OY 239 LeuGlyProAspGluGlnSerLeuHisProTrpGluGlnAspLeuLysIleCys 258
DB 919 CTCAAGCCCGGTAACCTCTCTCAAGCTGACGACGATCTTTTGACAAAGATTGT 978
OY 259 AspAlaTrpTrpLeuProAlaTrpCysSerThrSerAspTrpValLysLeuLeuIns 278
DB 979 GACCATTTCTACTGCTCCAGCTGCTGCTGCGCGCTGAGCTACGTCATGATGCCAAAC 1038
OY 279 LeuSerLeuGlnAspIleLysSerGluAspTrpSerArgPheValAlaProPheTrpPro 298
DB 1039 TTAGCGCTGCGAGATCAACAGGAGGCTGTGTCGAGTACGTCAGCCCTTCTGGCCA 1098
OY 299 AlaValIleArgSerAlaPheThrTrpLysGlyLeuSerSerLeuLeuSerGlyGln 318
DB 1099 GCGGTGATGAAACCGCTCTCTCATGTAAGGCTGTGGGACGTCGTCACGCTCGGCTCG 1158
OY 319 LysThrIleLysGlyAlaLeuAlaMetProLeuMetIleGluGlyTrpLysAspLeu 338
DB 1159 ACACATCTGAAGAGAGCTTTCGCTGACGCTCATGATCAGGCTACAGGCGGAGGCTC 1218
OY 339 IleLysPheAlaIleIleThrCysArgLys 348
DB 1219 ATTAATTCGCTGCATCACTTCAGAGAG 1248

RESULT 11
AAA29155
ID AAA29155 standard; cDNA; 1011 BP.
XX
AC AAA29155;
XX
DT 12-SEP-2000 (first entry)
XX
DE Wheat gamma tocopherol methyltransferase cDNA.
XX
KW Vitamin E: alpha-tocopherol; biosynthesis; enzyme; inhibitor; herbicide;
KW gamma-tocopherol methyltransferase; ss.
XX
OS Trilicium aestivum.
XX
XX Key Location/Qualifiers
XX CDS 3..885
XX FT /*tag- a
XX FT /product= gamma tocopherol methyltransferase
XX FT /transl_except= (pos:381..383, aa:Xaa)
XX FT /transl_except= (pos:387..389, aa:Xaa)
XX FT /transl_except= (pos:390..392, aa:Xaa)
XX FT /transl_except= (pos:393..395, aa:Xaa)

```

```

FT FT /transl_except= (pos:399..401, aa:Xaa)
XX XX /note= "Xaa is not defined"
PN WO200032757-A2.
XX
XX 08-JUN-2000.
XX
XX 02-DEC-1999; 99WO-US28588.
XX
XX 03-DEC-1998; 98US-0110781.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E. I.
XX
XX Cahoon RE, Coughlan SJ, Miao G, Rafalski JA;
XX
XX MPI: 2000-412309/35.
XX
XX P-PSDB: AAT96463.
XX
XX Polynucleotide encoding a vitamin E biosynthetic enzyme polypeptide
XX such as gamma tocopherol methyltransferase, useful for synthesizing
XX vitamin E or as a target for design and discovery of herbicides
XX
XX Claim 3; Page 53; 82pp; English.
XX
XX AAA29151-69 encode vitamin E (alpha-tocopherol) biosynthetic enzymes or
XX their fragments. The cDNA clones were identified by BLAST searches based
XX on similarity to Synecchocystis sp. and Arabidopsis thaliana vitamin E
XX biosynthetic enzymes. The enzymes are useful for synthesizing
XX plastoquinones or tocopherols (especially vitamin E). Vitamin E is
XX required for plant growth, therefore the enzymes may be used for the
XX discovery of new herbicides. The enzymes can be used in methods to
XX evaluate potential inhibitors, which may have use as herbicides.
XX Additionally, transgenic expression of, e.g. gamma-tocopherol
XX methyltransferase affords the ability to manipulate tocopherol levels as
XX desired for a particular application.
XX
XX Sequence 1011 BP; 255 A; 269 C; 268 G; 216 T; 3 other;
XX
XX Alignment Scores:
XX Pred. No.: 1.24e-89 Length: 1011
XX Score: 956.00 Matches: 194
XX Percent Similarity: 79.21% Conservative: 27
XX Best Local Similarity: 69.53% Mismatches: 55
XX Query Match: 52.24% Indels: 4
XX DB: 21 Gaps: 0

US-09-857-613A-28 (1-350) x AAA29155 (1-1011)
OY 75 GluSerSerGlyLeuTrpGluAsnIleTrpGlyAspHisMetHisGlyPheTrpAsp 94
DB 3 GACTCGTCCGCGCTGTGGAGAGCATCTGGGGAGACATGACACACGCGCTTCTACGAC 62
OY 95 SerAspSerThrValSerLeuSerAspHisArgAlaAlaGlnIleArgMetIleGlnGlu 114
DB 63 TCCGGGAGAGCCGCCCTCCATGTCGACACACCGCGCGGACGATGATGAGAG 122
OY 115 SerLeuArgPheAlaSerValSerGluGlnArgSerLysTrpProLysSerIleValAsp 134
DB 123 GCCCTGCTGCTGCGCGCTCCCGCGACATCGACAAACCAACCGATTTGTGAT 182
OY 135 ValGlyCysGlyIleGlyGlySerSerArgTrpLeu-AlaLysLysPheGlyAlaThrse 154
DB 183 GTTGATGCGGAGATGGGTGTAGCTCAATACCTGCGGCAACAATATGAGGACAATG 242
OY 134 rValGlyIleThrLeuSerProValGlnAlaGlnAlaAsnAlaLeuAlaAlaGln 174
DB 243 CTCGCGATCAACATTGAC-CCAGTCAAGCTGAGAGGAAATGCGCGCGCAGCCCA 301
OY 174 nGlyLeuAlaAspLysValSerPheGlnValAlaAspAlaLeuGlnGlnProPheSerAs 194
DB 302 AGGGGTTTCCGGAGCAAGGTTCTTCAATTGCTGATCTCGGAGCAACATTTCCTGG 361
OY 194 pGlyGlnPheAspLeuVal-TrpSerMetGluSerGlyGluHisMetProAspLysAla 214

```

```

Db 362 ATGGCGATTGATCTTCCGGGNCATGGAGAGANTGTGACACATCCGAGAACACAGA 421
Oy 214 yshvvalglglulealalargvalalalalproglalallellellevalthrt 234
Db 422 AGTTTGAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 481
Oy 234 rpyshlarsppluglulproasppluglulinsertleuhsiprotrppluglulinspleu 254
Db 482 GGTCCATAGACACCTCGCCGATCGGAGACTCAGTCAACCGACGACGACGACGACGACG 541
Oy 254 eulysllylecyasaplatatrryrlleuproalatrpyserthserapryval 274
Db 542 TGAATAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 601
Oy 274 yslleuenglinsertleuhsiproasppluglulinsertleuhsiproasppluglul 294
Db 602 AGATTCCGACGATCTCTTCTTGAAGATATCAAAAGCCGACGACGACGACGACGACG 661
Oy 294 larpophetrrproalalalalalalalalalalalalalalalalalalalalalal 721
Db 662 CCCCGTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 734
Oy 314 eusersergllylnlythrllelsglyalaleualametproleumetlelglulgt 334
Db 722 TAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 781
Oy 334 yrllyslasplleuilellysphe-Alallelelthrcysarglysprou 350
Db 782 ACAAGAAAGGCTCATTAAGTTCAAGCATCATCACTCCACAAACCCCA 832

RESULT 12
ID AAA29151 standard; cDNA: 792 BP.
AC AAA29151:
XX 12-SEP-2000 (first entry)
DE Corn gamma tocopherol methyltransferase partial coding sequence.
XX Vitamin E: alpha-tocopherol: biosynthesis; enzyme; inhibitor; herbicide;
KM gamma-tocopherol methyltransferase; ss.
XX Zea mays.
OS location/Qualifiers
FH Key 2..574
FT CDS /product= gamma_tocopherol_methyltransferase
FT /partial
XX WO200032757-A2.
XX 08-JUN-2000.
XX 02-DEC-1999; 99WO-US28588.
XX 03-DEC-1998; 98US-0110781.
XX (DUPO) DU PONT DE NEMOURS & CO E. I.
XX Cahoon RE, Coughlan SJ, Miao G, Rafalski JA:
XX WPI: 2000-412309/35.
XX P-PSDB: AAY96459.
XX polynucleotide encoding a vitamin E biosynthetic enzyme polypeptide
XX such as gamma tocopherol methyltransferase, useful for synthesizing
XX vitamin E or as a target for design and discovery of herbicides
XX Claim 3; Page 49; 82pp: English.

```

```

CC AAA29151-69 encode vitamin E (alpha-tocopherol) biosynthetic enzymes or
CC their fragments. The cDNA clones were identified by BLAST searches based
CC on similarity to Synecocystis sp. and Arabidopsis thaliana vitamin E
CC biosynthetic enzymes. The enzymes are useful for synthesizing
CC plastoquinones or tocopherols (especially vitamin E). Vitamin E is
CC required for plant growth, therefore the enzymes can be used in methods to
CC discovery of new herbicides. The enzymes can be used as herbicides.
CC evaluate potential inhibitors, which may have use as herbicides.
CC Additionally, transgene expression of, e.g. gamma-tocopherol
CC methyltransferase affords the ability to manipulate tocopherol levels as
CC desired for a particular application.
XX Sequence 792 BP; 201 A; 204 C; 215 G; 166 T; 6 other:
SO
Alignment Scores:
Pred. No.: 2,77e-71 Length: 792
Score: 778.00 Matches: 145
Percent Similarity: 85.86% Conservative: 19
Best Local Similarity: 75.92% Mismatches: 27
Query Match: 42.51% Indels: 0
DB: Gaps: 0
US-09-857-613a-28 (1-350) x AAA29151 (1-792)
Oy 159 leuserprovalglulalaglalalalalalalalalalalalalalalalalalal 178
Db 2 TTGAGCCCTGTCAAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 61
Oy 179 lysvalserpneglvalalalalalalalalalalalalalalalalalalalalal 198
Db 62 CAGGTTACTGTCAAGAGTGTCTGATGCTGAGGACCAACCGTTCTGAGGAGGAGGAGTTCAT 121
Oy 199 leuvaltrpsermetglulsergllylnlshmetproasppllyalalysphevalgl 218
Db 122 CTGGTGTGCTCATGAGAGGTGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 181
Oy 219 leuvalargvalalalalalalalalalalalalalalalalalalalalalal 238
Db 182 CTAGCAGCGGTGCGGCTCTCTGAGGACATATATCATGATGATGATGATGATGATGATG 241
Oy 239 leuglyproaspplulinsertleuhsiprotrppluglulinspleuilellylcy 258
Db 242 CTGATTCATCCAGAACCTGCTAAACCCGACGACGACGACGACGACGACGACGACGAC 301
Oy 259 aspalatrryrlleuproalatrpyserthserapryvallyslleuenglinsert 278
Db 302 GACGCTACTACTCCGAGACTGCTCTACCTTCAGACTATGTGACATTTGCCAGCTCA 361
Oy 279 leuserleuglinsplellyserglulasprrpserarqphevalalalalalalal 298
Db 362 CTGTCTCGAGATATCAAGACCTGCTGCGAGACGTCGCGCGCTTTTGCGCC 421
Oy 299 alavallleargseralaphethrrpplysgllyleuserleuserleuserleuser 318
Db 422 GCCGTGATTAATTCAGCCGCTACATGAGAGGCTTCACCTCTCTGCGAGCCGATGCG 481
Oy 319 lylthrllelsglyalaleualametproleumetlelglulgtlyrllyslaspleu 338
Db 482 AAGACGATCAGAGCCGCGATGATGATGATGATGATGATGATGATGATGATGATGATG 541
Oy 339 ilelsphealallelelthrcysarglysprou 349
Db 542 ATCAATTCACATCATCACCTGTCGCAAGCT 574

RESULT 13
ID AAX17788 standard; DNA: 954 BP.
AC AAX17788:
XX 21-MAY-1999 (first entry)
XX Synecocystis gamma-tocopherol methyltransferase encoding gene.

```


Key	Location/Qualifiers
FF	10..915
FT	/*cag= a
CT	/product= "MPWT"
TT	
YY	105
Db	217
YY	125
Db	274
YY	145
Db	331
YY	165
Db	391
YY	185
Db	451
YY	205
Db	511
YY	225
Db	571
YY	245
Db	631
YY	265
Db	691
YY	285
Db	751
YY	305
Db	811
YY	325
Db	871
YY	345
Db	931
RESULT 14	
ID	AAF26173
XX	AAF26173 standard; DNA; 930 BP.
XX	AAF26173:
DT	26-APR-2001 (first entry)
DE	Synechocystis FCC6803 MPWT encoding DNA SEQ ID 7.
XX	
KW	2-methyl-6-phytylhydrochinone methyltransferase; MPWT; transgenic plant
KW	2,3-diphosphol; tocopherol; vitamin E; 2,3-dimethyl-6-phytyl hydroquinone;
XX	2,3-dimethyl-6-geranylgeranyl hydroquinone; ds.
OS	Synechocystis sp.

[illegible]

XX Claim 2: Page 30-32; 34pp; English.

PS
CC The present sequence encodes a 2-methyl-6-phytylplastoquinol/2-methyl-6-solaniylplastoquinol-9 methyltransferase (methyltransferase 1) protein. The enzyme isolated from *Synechocystis* strain PCC6803. The enzyme is a fundamental enzyme in the production of tocopherols and plastoquinols in higher plants. The enzyme is believed to be involved in regulating the relative amounts of the various tocopherols in photosynthetic organisms. Genetic constructs comprising the methyltransferase 1 coding region under the control of a plant promoter are used to produce transgenic plants, to alter the ratio of delta-tocopherol:gamma-tocopherol in plants. This increases the nutritive value of the plants and products produced from them for humans and animals.

XX Sequence 957 BP; 203 A; 244 C; 268 G; 242 T; 0 other;

Alignment Scores:

Pred. No.:	4.7e-31	Length:	957
Score:	392.00	Matches:	92
Percent Similarity:	51.76%	Conservative:	40
Best Local Similarity:	36.08%	Mismatches:	87
Query Match:	21.42%	Indels:	36
DB:	21	Gaps:	10

US-09-857-613A-28 (1-350) x AA261599 (1-957)

OY 69 ILEAIGLPHETRYASPGLU-----SERSERGLYLEUTRPLUASNIETRPGLYASP 86
DB 106 GTGGCCACCGCTACGACCAATGACAGACAGACGATTTTGGAATATTACTGGCGCAC 165
OY 87 HSMETHSHISGLYPHETRYASPSERASPSERTHVALSERLEUSERASPHISARGALA 106
DB 166 CATATCCACCTGCGCATAT---GGCGATCCCGCCAGTGGCCAGAT-----210
OY 107 ALAIGLILEARGMETILEGLINUSERLEUARGPHEALASERVALSERGLUINARGSER 126
DB 211 -----TTCATCCATCGAATAATTGATTTCG-----CATGCCATGCC 249
OY 127 LYSTP-----PROLYSERILEVALASPVAILGYCYSGLY 138
DB 250 CAGTGGCGGCGATTAGTACTCTCCCGCCGACACAGGTATTGATGGGTGGCGC 309
OY 139 ILEGLYGLYSERASPTGYRLEUALALYSYPHEGLVALATHRSERVALGLYLETHR 158
DB 310 ATTGGCGGTACGACGCTTCTCCGCAAGATTATGTTTAACTTACCGGCATCACC 369
OY 159 LEUSERPROVALGLNALAGLINALAASNALALEUALALALAGLINGLYLEUALAASP 178
DB 370 ATTAGTCCCAACAGGTGAACGGCGGACGAGATTAACTCTCCGATGTGACGGCC---426
OY 179 LYSVALSERPHEGLINVALAASPALALEUGLINPROMPHESERASPGLYGINPHEASP 198
DB 427 -----AAGTTGCGGTGACGATGCTTGTCTTCTTCTTCTGACGCTACTTTCGAC 480
OY 199 LEUVALTRPSERMETGLUSERGLYUHSMEPROASPLYSALALYSYPHEVALGLYGLU 218
DB 481 GTAGTTGTGCTGCGTAGACAGGCCCCACATGCTGACAAAGCTGTGTTGCCAACGAA 540
OY 219 LEUALARGVALALALALAPROGLYALALEILEILEILEVALTHRTPCYSHISARGASP 238
DB 541 TTACTGCGGCGTGTGAACACGAGGCGCATTTCTGCTGCGCGGATTCGATCAACGGGAC 600
OY 239 LEUGLYPROASPLUGLN---SERLEUHSIPROTPRPLUGLINASPLEULEULYSYLE 257
DB 601 -----GATGCCCAAGTCCCTCAACTTCTGGAAAAACAGTAGCGCAACAACGTG 651
OY 258 CYSASPALATYRTRYLEUPROALATRPYSSERTHSERASPTRYVALYLSLEULEUGLN 277
DB 652 TTGGATCAATGGTCCCGCTCTTGGCAGCATTTGAAGGTTTGGGAAAAATTGGAA 711
OY 278 SERLEUSERLEUGLINASP-----ILEYSSERGLUASPTRPSERARGPHEVALALAPRO 295
DB 711

DB 712 GCCACGGGTTTGGTGGAGGGCCAGTGTACTACTGTGATTGACACTGTACCGACCTTCCCC 771
OY 296 PHEPTRPROVALVALLIARGSERALAPHEPTRIPLYSGLYLEU 310
DB 772 GCTTGGTTGGATACCATTT-----TGGCAGGGCATTT 801

Search completed: March 19, 2003, 07:20:17
Job time : 234 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 19, 2003, 07:11:35 ; Search time 49 Seconds
(without alignments)
2190.548 million cell updates/sec

Title: US-09-857-613A-28
Perfect score: 1830
Sequence: 1 MATVRIPTISCIHITFRS.....IEGYKDKLIFKAITCRKPE 350

Scoring table: BLOSUM62
Xgapop 10.0 , Ygapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 15338381 residues
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=x1h
-O=/cgn2_1/USPTO.spool/US09857613/runat_12032003_100037_6041/app_query.fasta_1.519
-DB=Issued_Patents.NA -QEMT=fastap -SUFFIX=rm1 -MINMATCH=0.1 -LOOPEL=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-USER=US09857613 -ECGN_1_1_32 -runat_12032003_100037_6041 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG-SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents.NA:*

- 1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PCUTS.COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	392	21.4	957	4	US-09-382-906A-1
2	261	14.3	2122	4	US-09-029-603A-1
3	225	12.3	6085	4	US-09-029-603A-4
4	220.5	12.0	80161	3	US-09-036-987A-1
5	220.5	12.0	80161	4	US-09-370-700-1
6	189	10.3	1400	4	US-09-041-718-1
7	185	10.1	49377	4	US-08-764-233A-1
8	156.5	8.6	1149	2	US-08-844-305-1
9	152.5	8.3	80161	3	US-09-036-987A-1
10	152.5	8.3	80161	4	US-09-370-700-1
11	138	7.5	4833	4	US-09-066-047-1
12	137.5	7.5	3236	4	US-08-961-527-222

13	131	7.2	8051	2	US-08-576-626A-2	Sequence 2, Appli
14	127	6.9	4403765	4	US-09-103-840A-2	Sequence 2, Appli
15	120	6.6	3719	1	US-08-920-812-10	Sequence 10, Appli
16	120	6.6	3719	1	US-08-920-827-10	Sequence 10, Appli
17	120	6.6	3719	1	US-08-921-177-10	Sequence 10, Appli
18	120	6.6	3719	1	US-08-362-577C-10	Sequence 10, Appli
19	120	6.6	3719	2	US-08-920-828-10	Sequence 10, Appli
20	117.5	6.4	4411529	4	US-09-103-840A-1	Sequence 1, Appli
21	114.5	6.3	2451	3	US-08-217-609A-21	Sequence 21, Appli
22	114.5	6.3	2451	4	US-08-873-235B-21	Sequence 21, Appli
23	114.5	6.3	4403765	4	US-09-103-840A-2	Sequence 2, Appli
24	113.5	6.2	741	3	US-09-217-609A-29	Sequence 29, Appli
25	113.5	6.2	741	4	US-08-873-235B-29	Sequence 29, Appli
26	109	6.0	9377	4	US-09-221-017B-1002	Sequence 1002, Ap
27	108.5	5.9	2077	4	US-09-221-017B-821	Sequence 821, App
28	108.5	5.9	4463	2	US-08-760-489-1	Sequence 1, Appli
29	108.5	5.9	4463	2	US-08-760-489-3	Sequence 3, Appli
30	108.5	5.9	4463	4	US-09-185-373-1	Sequence 1, Appli
31	108.5	5.9	4463	4	US-09-185-373-3	Sequence 3, Appli
32	107.5	5.9	8257	4	US-09-484-970B-65	Sequence 65, Appli
33	104	5.7	756	2	US-08-401-068-11	Sequence 11, Appli
34	104	5.7	756	2	US-08-846-338-11	Sequence 11, Appli
35	104	5.7	5872	3	US-08-411-768B-1	Sequence 1, Appli
36	104	5.7	5872	3	US-08-411-768B-6	Sequence 6, Appli
37	103	5.6	4911	4	US-09-718-692-1	Sequence 1, Appli
38	103	5.6	4911	4	US-09-718-852-1	Sequence 1, Appli
39	103	5.6	4911	4	US-09-718-815-1	Sequence 1, Appli
40	102.5	5.6	5100	1	US-08-457-245-1	Sequence 1, Appli
41	102.5	5.6	4411529	4	US-09-103-840A-1	Sequence 1, Appli
42	101.5	5.5	5822	3	US-08-899-595-4	Sequence 4, Appli
43	101.5	5.5	5822	3	US-08-899-595-5	Sequence 5, Appli
44	101	5.5	1406	4	US-08-936-165A-119	Sequence 119, App
45	99.5	5.4	777	4	US-09-134-001C-407	Sequence 407, App

ALIGNMENTS

RESULT 1
US-09-382-906A-1
Sequence 1, Application US/09382906A
Patent No. 6448475
GENERAL INFORMATION:
APPLICANT: Dellapenna, Dean
APPLICANT: Shintani, David
TITLE OF INVENTION: Manipulation of Tocopherol Levels in Transgenic Plants
FILE REFERENCE: 920905, 90032
CURRENT APPLICATION NUMBER: US/09/382, 906A
CURRENT FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 60/097, 863
PRIOR FILING DATE: 1998-08-25
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 957
TYPE: DNA
ORGANISM: Synechocystis PCC6803
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(954)
US-09-382-906A-1

ALIGNMENT Scores:

Pred. No.: 6,31e-37
Score: 392.00
Percent Similarity: 51.76%
Best Local Similarity: 36.08%
Query Match: 21.42%
DB: 4
Gaps: 10

US-09-857-613a-28 (1-350) x US-09-382-906A-1 (1-957)

QY 69 fileAclabpheryraspglu-----SerSerglyLeutrgcluasnlletrglyasp 86
:::||||| |||||::: ||| |||||

```
Db 106 GTGGCCACGCGCTACGACCAATGACAGAGCGCATTTTGGAAATATTCTGGGGCGAC 155
Qy 87 HisMetHisHisGlyPheTyraSerAspSerThrValSerLeuSerAspHisArgAla 106
Db 166 CATATCCACCTCGCCATTAT---GGCATCCCGCAGTGGCCAAAGAT----- 210
Qy 107 AlaGlnIleArgMetIleGlnIleuSerLeuArgPheAlaSerValSerGluIleuArgSer 126
Db 211 -----TTCAATCCAAATGCAAAATTGATTTGTC-----CATGCCATGCGC 249
Qy 127 LysTrp-----ProLysSerIleValAspValGlyGly 138
Db 250 CAGTGGGGCGGATTACATACACTTCCCGGACACAGGTATTTGATGTGGGTTCGGC 309
Qy 139 IleGlyIleSerSerArgTyrrLeuAlaLysLysPheGlyValaThrSerValGlyIleThr 158
Db 310 ATTAGGGGTAGAGTCGATTCCTCGCCAAAGATATGTTTAACTTACCGGATCACC 369
Qy 159 LeuSerProValGlnAlaGlnArgAlaAsnAlaLeuAlaIleGlnIleuAlaAsp 178
Db 370 ATTAGTCCCAACAGGTGAACCGGACGAGATTAATCTCCGATGTGACGGCC--- 426
Qy 179 LysValSerPheGlnValAlaAspAlaLeuGlnIleProPheSerAspGlyGlnPheAsp 198
Db 427 -----AAGTTTCGGGTGACAGCATGCTATGCTTCTTCTTCTGACGATTTTCGAC 480
Qy 199 LeuValTrpSerMetGlySerGlyIleuHisMetProAspLysAlaLysPheValGly 218
Db 481 GAGTTTGGTGGTGAACAGAGGCGCCCAATGCCATGCCAACAACCTGTGTTGCCAAGAA 540
Qy 219 LeuAlaArgValAlaAlaProGlyAlaIleIleIleValaThrTrpCysHisArgAsp 238
Db 541 TTACTCCGGGTGTGAACAGGCGGATCTGTGTGGTGGCGGATGCAATCAACCGGAC 600
Qy 239 LeuGlyProAspGluGln---SerLeuHisProTrpGlnIleuAspLeuLysIle 257
Db 601 -----GATCCCAATGCGCCCTCACTTCTGCGAAAACACGATGACCAACTG 651
Qy 258 CysAspAlaTrpTyrrLeuProAlaTrpCysSerThrSerAspTrpValLysLeuGln 277
Db 652 TTGGATCAATGGTCCACCCCTGCTTGGCAGCATGAGTTTCCGGAATTTGGAA 711
Qy 278 SerLeuSerLeuGlnAsp-----IleLysSerGluAspTrpSerArgPheValAlaPro 295
Db 712 GCCAGCGGTTGTGGAGGCGCAGGACTACTGCTGATTGGACGTACCGACCCCTCC 771
Qy 296 PheTrpProAlaValIleArgSerAlaPheThrTrpLysGlyLeu 310
Db 772 GCTTGGTTGATNCCATT-----TGGCAGGCGATT 801

RESULT 2
US-09-029-603-1
: Sequence 1, Application US/09029603
: Patent No. 6210935
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Engel, Natalie
: APPLICANT: Bietenhader, Jurg
: APPLICANT: Toupet, Christine
: APPLICANT: Pospisich, Andreas
: TITLE OF INVENTION: Staurosporin Biosynthesis Gene Clusters
: FILE REFERENCE: 4-20355/A/PCT
: CURRENT APPLICATION NUMBER: US/09/029, 603
: EARLIER FILING DATE: 1998-03-20
: EARLIER APPLICATION NUMBER: PCT/EP96/03643
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 2122
: TYPE: DNA
: ORGANISM: Streptomyces longisporoflavus
: FEATURE:
```

```
: NAME/KEY: misc_RNA
: LOCATION: (1)..(2122)
: OTHER INFORMATION: product = 2.1 kb region
US-09-029-603-1

Alignment Scores:
Pred. No.: 7, 81e-21
Score: 261.00
Percent Similarity: 46.59%
Best Local Similarity: 28.67%
Query Match: 14.268
DB: 4 Gaps: 10

US-09-857-613a-28 (1-350) x US-09-029-603-1 (1-2122)

Qy 19 ArgSerGlnSerProArgThrPheAlaArgIleArgValGlyProArgSerTrpAlaPro 38
Db 733 CGGGAGCTGCTCCCGCCG-----GCGAGCCCGCGCGCGTACCGCCGACCGACGACGA 786
Qy 39 IleArgAlaSerAlaIleAspSerGluArgGlyIleIleValLeuGlnIleuLysProLys 58
Db 787 TTTCACCTCCCTTCTACCTCCCTTCTCTCACTGATCGATCCGAAAGCCGTTCCCATG 846
Qy 59 LysAspAspLysLysLysLeu-----GlnLysGlyIleAlaGluPheTyraSpGlu--- 75
Db 847 ACCGATCCACACCCAGACCTGCGCCGCGGAGAGCGGTGCTGATGCTGATGACCGCGCTG 906
Qy 76 ---SerSerGlyLeuTrpGluAsnIleTrpGlyAspHisMetHisGlyPheTyraSp 94
Db 907 ACCTGAGCGCGATGACAGCAGCGCTGTCACCCCAATGTCGACATCGGCTTTGGGAC 966
Qy 95 -----SerAspSerThrValSerLeuSerAspHisArgAlaIleArgMet 111
Db 967 ACCCGCGGCTCGAGCGCCACCATCGAGAGGCGATGAGCGGCTGACCGATGTC---TTC 1023
Qy 112 IleGlnIleSerLeuArgPheAlaSerValSerGluIleuArgSerLysTrpProLysSer 131
Db 1024 ATCGAAGCGCTGACCGCGTACCGCCAC-----TCCAC 1056
Qy 132 IleValAspValGlyCysGlyIleGlyLysSerSerArgTyrrLeuAlaLysPheGly 151
Db 1057 GTCCTGACCTGCGGTGGGGGGGGGGGGGGGGGGCTGAGGTCTGTGGCGCCACCGGG 1116
Qy 152 AlaThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaAsnAlaLeuAla 171
Db 1117 GCACGGGTACCGCGCATACGATCAGCAGAGACGACATAGACCGCCCAACCGGCTGGCC 1176
Qy 172 AlaAlaGlnGlyLeuAlaAspLysValSerPheGlnValAlaAspAlaLeuGlnIlePro 191
Db 1177 GCCGAGCGCGGGGTGCGCAGCGCTGCTTCACCATGCGCAGCGCATGAAATGCGCC 1236
Qy 192 PheSerAspGlyGlnPheAspLeuValTrpSerMetGluSerGlyIleuHisMetProAsp 211
Db 1237 TTGCGCGAGCGCTGCTGACCGCCGATGAGCGGTGAGTGTGTCACCATGACCGCCGAC 1296
Qy 212 LysAlaLysPheValGlyLysLeuAlaArgValAlaAlaIleProGlyAlaIleIleIle 231
Db 1297 CGGACGAGGTGTCACCGAGGTGCGCGGTGCTGCGCCCGGCGGCGGATGTCCTC 1356
Qy 232 ValThrTrpCysHisArgAspLeuGlyProAspGluGlnIleuHisPro----- 248
Db 1357 ACCGACATCTGAGAGGCGC-----CCGCGCAAGCGCGTACGACACCCCGCATCGAC 1410
Qy 249 ---TrpGluGlnAspLeuLysLysIleCysAspAlaTrpTyrrLeuProAlaTrpCys 267
Db 1411 AAGTCTGCGCGGACTGTATGTCGACACGCGGAC----- 1446
Qy 268 SerThrSerAspTrpValLysLeu-----GlnSerLeuSerLeuGlnAspIle 284
Db 1447 ---ATCGAGCATCTACGTGCGCTGCTGACCGCGCTCCGGGCTGCGGTGCGGAGATC 1500

RESULT 3
US-09-029-603-4
```

```

: Sequence 4, Application US/09029603
: Patent No. 6210935
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Engel, Natalie
: APPLICANT: Bietenhader, Jurg
: APPLICANT: Toupet, Christine
: APPLICANT: Pospiech, Andreas
: TITLE OF INVENTION: Staurosporin Biosynthesis Gene Clusters
: FILE REFERENCE: 4-20555/A/PCT
: CURRENT APPLICATION NUMBER: US/09/029,603
: EARLIER FILING DATE: 1998-03-20
: EARLIER APPLICATION NUMBER: PCT/EP96/03643
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4
: LENGTH: 6085
: TYPE: DNA
: ORGANISM: Streptomyces longisporoflavus
: FEATURE:
: NAME/KEY: misc_RNA
: LOCATION: (378)..(1665)
: OTHER INFORMATION: ORF
: FEATURE:
: NAME/KEY: misc_RNA
: LOCATION: (1747)..(2553)
: OTHER INFORMATION: ORF
: FEATURE:
: NAME/KEY: misc_RNA
: LOCATION: (2593)..(4011)
: OTHER INFORMATION: ORF
: FEATURE:
: NAME/KEY: misc_RNA
: LOCATION: (4013)..(4999)
: OTHER INFORMATION: ORF
: NAME/KEY:
: LOCATION: (5071)..(6085)
: OTHER INFORMATION: ORF
: US-09-029-603-4

Alignment Scores:
Pred. No.: 7.93e-16 Length: 6085
Score: 225.00 Matches: 75
Percent Similarity: 41.458 Conservative: 51
Best Local Similarity: 24.678 Mismatches: 120
Query Match: 12.30% Indels: 58
DB: Gaps: 12

US-09-857-613a-28 (1-350) x US-09-029-603-4 (1-6085)
QY 33 ProaHSeTTPAlaProIleArgAlaSerAlaSerSer-----GIuaTgLyGlu 50
DB 1684 CCCTTCCTCCCTGGCCACCCCTCGACCCCTCTCGACTCGACGACGAGGAGGACACA 1743
QY 51 lIleValleuGluInLysProLysLysAspAspLysLysLysLeuGluInLysGlylIleAla 70
DB 1744 TCCATACGACGAGTCGACGACACCCGCGGAC-----TCGGTCGGT 1785
QY 71 GIuPHeTyTAspLlu-----SerSerGlyLeuTTPcluaSnIlleTTPclYAspHisMet 88
DB 1786 GAGGTCTAGACACAGTTGCCGCGGCGGCGCC---AGCACCGCGAGTGGCGGCGACACATC 1842
QY 89 HIsHISGlyPHeTyTAspSerAspSerThnValSerLeuSerAspHisArgAlaAlaGln 108
DB 1843 CACGTGGGTAATCGGACGACGACCCGAGGTGCGATCGCGAGGACGACGCGGCTC 1902
QY 109 lIleArgMetlIeGluInLysSerLeuArgPheAlaSerValSerLcIuArgSerLysTTP 128
DB 1903 ACCGATCTCTCGCCGAGCGCGCTCGGCGCTC-----CGCCCG 1938
QY 129 ProLysSerlIeValAspValGlyGlylIleGlyLysSerArgTgTLeuAlaLys 148

```

```

DB 1939 GACCGCATCTGCTGACCTGGGCTCGGCATCGCGCTCCGCTCTCAGATCCGCGGA 1998
QY 149 LysPHeGlyAlaThrsSerValGlylIeThrsSerProValGlnAlaInLrArgAlaSn 168
DB 1999 GCGCACGACGCTCCGCTCACCGGATCACGATCAGCACGACGATCACCGAGCGGCC 2058
QY 169 AlAlaValAlaAlaGlnGlyLeuAlaAspLysValSerPHeGlnValAlaAspAlaLeu 188
DB 2059 GAGCGGCGGTGAGTCGATCGATCCGCGGCGCGGTCTCTCCGCTGGCGGCGGCAATC 2118
QY 189 GlnGlnProPHeSerAspGlyGlnPheAspLeuValTTPSerMetLcLysSerGlyLHis 208
DB 2119 GACCTCCCTTCGAGGACGCTCTCTCGACGCGCGCTTCGATCAGTCGCTGCTCAT 2178
QY 209 MetProAspLysAlaLysPheValGlyGluLeuAlaArgValAlaAlaProGlyAlaLe 228
DB 2179 CTGCGCGACGACGACCGCGCGCTCAAGAGATCCACCGGCTGCTCCGCCCGGCGCG 2238
QY 229 lIeIlelIleValThrTPcysHisArgAsp-----LeuGlyProAspLcLysSerLeu 246
DB 2239 CTGCTATCGCGACCTGTGTACGACGACGCGCTTACCGCGCGCGGACGAGGTGCTC 2298
QY 247 His-----ProTTPclLcLysSerLeu 253
DB 2299 GACGCGATCTGCTGATGTACGAGATCGCGGATCAGACACACCTACGACGAT----- 2352
QY 254 LeuLysLyslIeCysAspAlaTgTTPLeuProAlaTTPcysSerThrsSerAspTgVal 273
DB 2353 CGCGCGGACGCTGGGAGGCG-----GGCTGG----- 2379
QY 274 LysLeuLeuGlnSerLeuSerLeuGlnAspLleLysSerGlnAspTTPSerArgPheVal 293
DB 2380 -----GACTGCTGAGCTGACGACGACATCGGAGACAG-----GTC 2415
QY 294 AlaProPHeTTPProAlaValIleArgSerAlaPheThrTTPcysLysLeuSerSerLeu 313
DB 2416 CGCGCTACTACGCG-----CATGCCGCGCGCGCTTCGCGGCGTCTCGCGGCGCT 2466
QY 314 LeuSerSerGly 317
DB 2467 CTCGACGCGCGC 2478

RESULT 4
US-09-036-987A-1
: Sequence 1, Application US/09036987A
: Patent No. 6143526
: GENERAL INFORMATION:
: APPLICANT: Baltz, Richard H.
: APPLICANT: Broughton, Mary C.
: APPLICANT: Crawford, Kathryn P.
: APPLICANT: Madduri, Krishnamurthy
: APPLICANT: Merlo, Donald J.
: APPLICANT: Treadway, Patli J.
: APPLICANT: Turner, Jan R.
: APPLICANT: Waldron, Clive
: TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
: NUMBER OF SEQUENCES: 39
: CORRESPONDENCE ADDRESS:
: STREET: 9330 Zionsville Road
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: USA
: ZIP: 46268
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/036,987A

```

```

: FILING DATE: 09-MAR-1998.
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Stuart, Donald R.
: REGISTRATION NUMBER: 28,479
: REFERENCE/DOCKET NUMBER: 50,608
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (317)337-4816
: TELEFAX: (317)337-4847
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 80161 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-09-036-987A-1

Alignment Scores:
Pred. No.: 1.73e-13 Length: 80161
Score: 220.50 Matches: 78
Percent Similarity: 43.99% Conservative: 50
Best Local Similarity: 26.80% Mismatches: 125
Query Match: 12.05% Indels: 39
DB: 3 Gaps: 9

US-09-857-613A-28 (1-350) x US-09-036-987A-1 (1-80161)
Oy 69 lleaagluhpheryraspluserSerglyLeutrpGlusnllertpGlyasp----- 86
Db 20201 GTTGGGCGAGATGATGACGTGTCACGCGCTGTCGACACTCGCTCCGCGCGCCCTGC 20260
Oy 87 HSMethSHISglYpHeryraspserserThValserleuserasphISargala 106
Db 20261 GCCATCCACGAGGACTGAGGAGACGAGCGGGCGCTTCCTGGACAGAGC-CCGCCA 20319
Oy 107 AlaglnleargmetlleglnluserleuargPhealaservalsergluGlyargser 126
Db 20320 CCGGCTCACCGACCTGTGCGCGAAGC-----GACCTGTCT 20355
Oy 127 LysTrpProLysSer-lleValasPvalglYcsglylleglylYserSeraRgTyrla 146
Db 20356 CGATGGCGGCGTTCGACTGCTGATGTGGGTGCGGTACCGCAACAGCGCTGCGCGT 20415
Oy 146 uAlaLysLysPheglYAlaThrservalGlylertHrleuserProValGlnAlaGlnr 166
Db 20416 CCGGCGCGCAACGCGATCCAGATCACCGCATCCGCGTACGCGAGTGGAGTGGCCAT 20475
Oy 166 gAlaAsnaLaLeuAlaAlaAlaGlnGlyLeuAlaAsPlysValserPheGlnValAlaAs 186
Db 20476 CCGCCCTGATTCGCGACCGGAGACCGGAGTAAACCGCGGTGATTCCTCGCGCTCA 20535
Oy 186 pAlaLeuGlnGlnProPheSeraSpIcYlnPheasPleuValTrpSerMetGlnSergI 206
Db 20536 TTCATGTCCTCCGCTCCGCGGACATGCTTCGACGCGCGCTGGGCGCATGACGTCGCT 20595
Oy 206 yLunHSMetProasPlysAlaLysPheValGlyGlnLeuAlaArgValAlaAlaProG1 226
Db 20596 GTTGGAGATGTCGGAACCGGACCGCTCCATCCGGAATCTCTTGAGTACTCAACCCGG 20655
Oy 226 yAlaAlaIlellelleValAlaThrTrpYshIasrAsPleuLyProasp---GlnGlnSe 245
Db 20656 TGGCATCTCTCGGCGTACCGAGGTCTCTCAACGAGAAGCGGGGGGATGTCGGCTGC 20715
Oy 245 rLeuHISProTrpGlnGlnAsPleuLeuLysLysIleCysAsPalaTyTrtYrleuProal 265
Db 20716 CCGGAGACAGTGGCGACCGGCTT-----CGATCTGCTGCGCT----- 20755
Oy 265 aTrpCysSerThrsasPtyrValLysLeuLeuGlnSerleuserleuGlnAsPleuLy 285
Db 20756 -----GAGCAACTTGTGAGATCGCTCGTCCACCGGGGTTTCA 20793
Oy 285 sSergLusAsPTrp-----SeraRgPheValAlaProPheTrpProal 299

```

```

Db 20794 GATCCCTCATTTGGAGAGACGCTGCTCGAGAGACCCGCGTCTTCAGCCGCACTTCCGCA 20853
Oy 299 aValleatYserAlaPheThrTrpLysGlyLeuserleuserSergIyGlnly 319
Db 20854 AGAGCTC-----GCTGGCGCACACCGGATTCGCGGACAGTACGGG-----CC 20898
Oy 319 sThrleLysGlyAlaLeuAlaMetProleuMetlleglYlYrTyLysAsPleu1 339
Db 20899 GCGTCTGCGCGCTGCGCGCGCGCGCTGCGCATTAATGAAATATGCCACAGACATG-- 20956
Oy 339 eLysPheAlaIlelleThrCysArgLysPro 349
Db 20957 -GGTATGCGATTGTACGCGCGGAGACCG 20986

RESULT 5
US-09-370-700-1
: Sequence 1, Application US/09370700
: Patent No. 6274350
: GENERAL INFORMATION:
: APPLICANT: Baltz, Richard H
: APPLICANT: Broughton, Mary C
: APPLICANT: Crawford, Kathryn P
: APPLICANT: Madduri, Krishnamurthy
: APPLICANT: Treadway, Patli J
: APPLICANT: Turner, Jan R
: APPLICANT: Waldron, Clive
: TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
: FILE REFERENCE: 50489 DIV1
: CURRENT APPLICATION NUMBER: US/09/370,700
: EARLIER FILING DATE: 1999-08-09
: EARLIER APPLICATION NUMBER: US 09/36987
: NUMBER OF SEQ ID NOS: 39
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 80161
: TYPE: DNA
: ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Alignment Scores:
Pred. No.: 1.73e-13 Length: 80161
Score: 220.50 Matches: 78
Percent Similarity: 43.99% Conservative: 50
Best Local Similarity: 26.80% Mismatches: 125
Query Match: 12.05% Indels: 39
DB: 4 Gaps: 9

US-09-857-613A-28 (1-350) x US-09-370-700-1 (1-80161)
Oy 69 lleaagluhpheryraspluserSerglyLeutrpGlusnllertpGlyasp----- 86
Db 20201 GTTGGGCGAGATGATGACGTGTCACGCGCTGTCGACACTCGCTCCGCGCGCCCTGC 20260
Oy 87 HSMethSHISglYpHeryraspserserThValserleuserasphISargala 106
Db 20261 GCCATCCACGAGGACTGAGGAGACGAGCGGGCGCTTCCTGGACAGAGC-CCGCCA 20319
Oy 107 AlaglnleargmetlleglnluserleuargPhealaservalsergluGlyargser 126
Db 20320 CCGGCTCACCGACCTGTGCGCGAAGC-----GACCGTGTCT 20355
Oy 127 LysTrpProLysSer-lleValasPvalglYcsglylleglylYserSeraRgTyrla 146
Db 20356 CGATGGCGGCGTTCGACTGCTGATGTGGGTGCGGTACCGCAACAGCGCTGCGCGT 20415
Oy 146 uAlaLysLysPheglYAlaThrservalGlylertHrleuserProValGlnAlaGlnr 166
Db 20416 CCGGCGCGCAACGCGATCCAGATCACCGCATCCGCGTACGCGAGTGGAGTGGCCAT 20475
Oy 166 gAlaAsnaLaLeuAlaAlaAlaGlnGlyLeuAlaAsPlysValserPheGlnValAlaAs 186

```



```

Db      1198 GGTGGTAGCAAAATAATTGTTACTCCATGCATGCATGTTAGCTTGTAACCAA 1251
          |||           |::|||           |::|||           |||||||
          332 GlucyltyrlylsaspIlelleysphalallethrcysArgLysPro 349
          Yy              ::::Yy              ::::Yy              Yy
RESULT 7
05-08-764-233A-1
: Sequence 1, Application US/08764233A
: Patent No. 5716849
GENERAL INFORMATION:
APPLICANT: Ligon, James M.
APPLICANT: Schupp, Thomas
APPLICANT: Beck, James J.
APPLICANT: Hill, Dwight S.
APPLICANT: Neff, Snezana
APPLICANT: Ryals, John A.
TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,233A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/729,214
FILING DATE: 09-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Melgs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: 1506/CIP6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 49377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Sorangium cellulosum
IMMEDIATE SOURCE:
CLONE: p98/1, pUJ3, and pVKM15
FEATURE:
NAME/KEY: misc.feature
LOCATION: 383..760
OTHER INFORMATION: /product= "Sorr"
/note= "this gene encodes a protein that is highly homologous
to type I PKs such as eryA from
other information:"
OTHER INFORMATION: Saccharopolyspora erythraea."
FEATURE:
NAME/KEY: misc.feature
LOCATION: 927..19874
OTHER INFORMATION: /product= "SorA"
/note= "Gene product is highly homologous to type I PKs that
are known to be involved in the synthesis of polyketide
compounds."
OTHER INFORMATION: compounds."
FEATURE:
NAME/KEY: misc_feature
```

```

LOCATION: 942..7115
OTHER INFORMATION: /product= "Module 1 of SorA"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7203..12884
OTHER INFORMATION: /product= "Module 2 of SorA"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 13455..19616
OTHER INFORMATION: /product= "Module 3 of SorA"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19870..24556
OTHER INFORMATION: /product= "Module 1 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 24638..30820
OTHER INFORMATION: /product= "Module 2 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 30881..35446
OTHER INFORMATION: /product= "Module 3 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 35528..40114
OTHER INFORMATION: /product= "Module 4 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 40190..46318
OTHER INFORMATION: /product= "Module 5 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 46851..47891
OTHER INFORMATION: /product= "SORM"
OTHER INFORMATION: /note= "The protein encoded by the sorm gene is highly homologous to the methyltransferase from Streptomyces OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the OTHER INFORMATION: polyketide rapamycin."
US-08-764-233A-1
Alignment Scores:
Prid. No.: 1.28e-09 Length: 49377
Score: 185.00 Matches: 89
Percent Similarity: 39.788 Conservative: 57
Best Local Similarity: 24.258 Mismatches: 129
Query Match: 10.11% Indels: 92
DB: 1 Gaps: 16
US-09-857-613A-28 (1-350) x US-08-764-233A-1 (1-49377)
QY 17 ThrPheargSerGlnserProargThrPhealaArgIleargValGlyProargserTrp 36
:::|||||::: ||| |||
Db 46821 AGCTCCGCGGGGCGATGAGGACGCCCATGCAAAATTTCGCCAGGGGCTCGGACA--- 46877
:::|||||::: ||| |||
QY 37 AlaProtle-----ArgAlaserAlaIaaserSerGluargGlyIle 51
:::|||||::: ||| |||
Db 46878 TCTCCCATCTCAACGCGATGTGTGAGCCGATGGCCAGCATGTCCAGACAGATCGAGCTG 46937
:::|||||::: ||| |||
QY 52 ValIeugIuGlnIysProIysLysaspIysLysIysLeuGlnIysely----- 68
:::|||||::: ||| |||
Db 46938 ATCAAAATGTTCAACGCGATGTGTGAGCCGATGGCCAGCATGTCCAGACAGCTG 46997
:::|||||::: ||| |||
QY 69 -----IleAlaGluPheTyasp--- 74
:::|||||::: ||| |||
Db 46998 GACGCGCTCGCCGCTCAAGGTGGACGCTCAAGCCCTATGAGACGCGCTCAACCTC 47057
:::|||||::: ||| |||
QY 75 -----GluSerSerGlyLeuTrpGluAsnIleTrpGlyAspHisMetHisGly 91
:::|||||::: ||| |||

```

```

Db 47058 GAGCCGCTCAAGTCGCCGATCCTTCACAAAC---TGGCGG-----ATGTTTCAGCAG 47108
Qy 92 PheTYrAspSerSerThrValSerLeuSerAspHisArgAlaIleIleArgMet 111
Db 47109 GAGGTTTTCGTGATGTTCTTCACACAGCTGCCAGATTACGACAAAGCAGCAGCAGCAGT 47168
Qy 112 IlegInlu-----SerLeuArgPheAlaSerValSerGluGluArgSer 126
Db 47169 TATTGAGACAGCTCTACGCTACACGCTGAAGCAAGTTCCGGCCGAGCAAGATCGG--- 47225
Qy 127 LysTrpProLysSerIleValAlaAspValGlyCysGlyIleGlyGlySerSerArgTrpLeu 146
Db 47226 -----CCACCCAGATCTCTGAGCTCGCGCAACACTGCGAGAGCGCTCAACTCTCT 47279
Qy 147 AlaLysLysPheGly---AlaThrSerValGlyIleThrLeuSerProValGlnIleArg 165
Db 47280 TCCCGATCGAGGCTCGCAGACACCTTCGCGCTCGACCTGCTCCAGCAGCGGTGCAGC 47339
Qy 166 ArgAlaAsnAlaLeuAlaAlaIleGlnIleLeuAlaAspLysValSerPheGlnValAla 185
Db 47340 ATCCGCAACGCCCGGCTTCGCGACCGGGC-----TCGTTGACCTGACCTTCAGGC 47390
Qy 186 AspAlaLeuGlnIleProPheSerAspGlyGlnPheAspLeuValTrpSerMetGluSer 205
Db 47391 GACCGGAGAACTCTTCCTTCGCGCAGCGCAATTCGACGCTCATCATCATTCGAGAGC 47450
Qy 206 GlyLysHisMetProAspLysAlaLysPheValGlyGluLeuAlaArgValAlaIlePro 225
Db 47451 TCGCAACACTACCTCAATCTCGCAATCATCTCGAAGTGGCGGCTGCTGCGACCG 47510
Qy 226 GlyAlaIleIleIleIleValThrTrpCysHisArgAspLeuGlyProAspGluIleSer 245
Db 47511 GCGCGGCTC-----TTTCGATGTGATGTCTCTCGCAACATGCGC--- 47552
Qy 246 LeuHisProTrpGluGlnAspLeuLeuLysIleCysAspAlaTrpTrpLeuProAla 265
Db 47553 -----TATTGGTGTATGCGAG 47567
Qy 266 TrpCys-----SerThrSerAspTrpValLysLeuGlnSerLeuSerLeuGlnAsp 283
Db 47568 AATTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 47627
Qy 284 IleLysSerGluAspTrpSerArgPheValAlaProPheTrpProAlaValIleArgSer 303
Db 47628 TACGTAAGAAGAGCTATCCGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 47666
Qy 304 AlaPheTrpTrpLysGlyLeuSerSerLeuLeuSerSerGlyGlnLysThrIleLysGly 323
Db 47667 -----ACCAAGCGCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 47693
Qy 324 AlaLeuAlaMetProLeu-----MetIleGlnGlyTrpTrpLys 336
Db 47694 GCGCTGCGCTACCCGCTCGGAAGCTTTTCAGCTCGGTGATGATGATGCGGCTACGAGATCC 47753
Qy 337 AspLeuIleLysPheAlaIle 343
Db 47754 GAG-----TTTGGCGTC 47765

RESULT 8
US-08-844-305-1
; Sequence 1, Application US/08844305
; Patent No. 5936139
; GENERAL INFORMATION:
; APPLICANT: Schmid, Katherine M.
; TITLE OF INVENTION: Cyclopropane Fatty Acid Expression in Plants
; FILE REFERENCE: MFS-403F
; CURRENT APPLICATION NUMBER: US/08/844,305
; CURRENT FILING DATE: 1997-04-10
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1149
; TYPE: DNA

```

```

; ORGANISM: Escherichia coli
US-08-844-305-1

Alignment Scores:
Pred. No.: 7,18e-09 Length: 1149
Score: 156.50 Matches: 69
Percent Similarity: 41.45% Conservative: 45
Best Local Similarity: 25.09% Mismatches: 105
Query Match: 8.55% Indels: 56
DB: 2 Gaps: 14

US-09-857-613a-28 (1-350) x US-08-844-305-1 (1-1149)

Qy 71 GluPheTYrAspGluSerSerGlyLeuTrpGluAsnIleTrpGlyAspHisMetHis 90
Db 352 GAGCATTTACGATTTGGGTATGACTTGTTCACCGCATCTTGATCCCTTCATGCAATAT 411
Qy 91 Gly-----PheTYrAspSerSerThrValSerLeuSerAspHisArgAlaAla 107
Db 412 TCTCGCGCTTACTGCAAGATGCCGAT-----AATCTGGAATCTGCCAGCAGCGC 462
Qy 108 GlnIleArgMetIleGlnIleSerLeuArgPheAlaSerValSerGluGluArgSerLys 127
Db 463 AAGCTCAAAATGATTTGTAAAAATTGCAGTTA-----AAA 498
Qy 128 TrpProLysSerIleValAspValGlyCysGlyIleGlySerSerArgTrpLeuAla 147
Db 499 CCAGGATCGCGCTACGATGATATGCTGCTGCGCTGGCGCGGACTGCACTACATGATGCA 558
Qy 148 LysLysPheGlyAlaThrSerValGlyIleThrLeuSerProValGlnIleArgAla 167
Db 559 TCTAATATAGACGTAGACCTGCGCGCGCTCACCATTTCGCCAGACGAAAAAATGCGCT 618
Qy 168 AsnAlaLeuAlaAlaIleGlnIleLeuAlaAspLysValSerPheGlnValAlaAspAla 187
Db 619 CAG-----GAACGCTGTAGAGCCTGATGTCCATT--- 651
Qy 188 LeuGlnIleProPheSerAsp-----GlyGlnPheAspLeuValTrpSerMetGluSer 205
Db 652 TTGCTCCAAAGATATGCTGATGACCGCAGACGACAGCTTGATGATGCTTCTGCGGAG 711
Qy 206 GlyGlnHisMetProAspLys-----AlaLysPheValGlyGluLeuAlaArgValAla 223
Db 712 TTGACAGCAGCTGCGAGCAGCAAAATTCAGATACCTATTGCGGTGATGATGATTTG 771
Qy 224 AlaProGlyAlaIleIleIleIleValThrTrpCysHisArgAspLeuLysProAspGly 243
Db 772 AAACCGAAGGCAATATTCCTGCTCCATCT-----ATCGTTCCGAAAAA 816
Qy 244 GlnSerLeuHis-----ProTrpGluGlnAspLeuLeuLysIleCysAspAlaTrp 261
Db 817 ACCGATCTAATATGATGATCCCTG-----ATTATTAATAT 852
Qy 262 TYrLeuProAlaTrpCysSerThrSerAspTrpValLysLeuGlnSerLeuSerLeu 281
Db 853 ATTTTCCGAGCGCTGCGCCCTCT-----GTACGCCAGATTCTCAGTCAGCGCAA 906
Qy 282 GlnAspIleLysSerGluAspTrpSerArgPheValAlaProPheTrpProAlaValIle 301
Db 907 CCCCACTTTGTGATGAGACACTGCACTTAACCTTCGCTGATTAAC----- 951
Qy 302 ArgSerAlaPheThr--TrpLysGlyLeuSerSerLeuLeuSerSerGlyGlnLys----- 319
Db 952 ----GATACTAAGTTGATGGGTGTGTGAGCAATTCCTGCCCGCATGSCAGAAATTGCG 1008
Qy 320 ----ThrIleLysGlyAlaLeu-----AlaMetProLeuMet 330
Db 1009 GATAACTATAGTGAACGCTTTAAACGAATGTTTACTATTATTC 1051

RESULT 9
US-09-036-987A-1/c
; Sequence 1, Application US/09036987A
; Patent No. 6143526

```

```

GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patli J.
APPLICANT: Turner, Jan R.
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
TITLE OF INVENTION: Production
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R.
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 80161 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1

Alignment Scores:
Pred. No.: 2e-05 Length: 80161
Score: 152.50 Matches: 48
Percent Similarity: 38.38% Conservative: 28
Best Local Similarity: 24.24% Mismatches: 77
Query Match: 8.33% Indels: 45
Gaps: 4
DB: 3

US-09-857-613A-28 (1-350) x US-09-036-987A-1 (1-80161)

QY 85 GlyAspHisMetHisGlyPheTyrAsp----- 94
DB 13517 GGGCGTCCCTCCACCGGCTACTGGGCGGGGTATCGGAGATCCCGTCCACA 13458
QY 95 -----SerAspSerThrValSerLeuSerAspHisArgAlaAlaGlnIleArgMetIle 112
DB 13457 CCGTGTCCGATGCTCCGCCACCACTGACCGAC-----CTGTTCACTGAC 13413
QY 113 GlnGluSerLeuArgPheAlaSerValSerGluGluArgSerIleTyrProIleSerIle 132
DB 13412 AAGGCCCGCGCTCCGCCGAGCG-----CACCTG 13383
QY 133 ValAspValGlyGlySerGlyIleGlyGlySerArgTyrLeuAlaLysLysPheGlyAla 152
DB 13382 TTCGACTGCGGCTGGCGCATGGGCGAGCCCGTAGTCCGTGGCGATGCCCGCGGCTT 13323
QY 153 ThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaAsnAlaLeuAla 172
DB 13322 CGAGTCAACCGGATCACTGCTGAACGCCGACGACATCTCGCGCGCCACGAGGCTCGCAAC 13263

```

```

QY 173 AlaGlnGlyLeuAlaAspLysValSerPheGlnValAlaAspAlaLeuGlnGlnPhe 192
DB 13262 GAGACCGGACTGGCCCGGACATCTTGACTGATCTGACAGCGCGCCACTGCTCCAC 13203
QY 193 SerAspGlyGlnPheAspLeuValTyrPheSerMetCysLeuSerGlyGlnHisMetProAspLys 212
DB 13202 CCGGACGCTTCTTTCAGCGCGCATGGCGCATGCTCGTGCAGATCGAGTGGACAG 13143
QY 213 AlaLysPheValGlyGlnLeuAlaArgValAlaAlaLapProGlyAlaIleIleIleVal 232
DB 13142 GCCGCCGCGATCCGCGAGGTCCACCGAATCTCGAACCAGCGCGCGGCTTGCTCCGA 13083
QY 233 ThrTyrPheHisArgAspLeuGlyProAspGluGlnSerLeuHisProTyrGluGlnAsp 252
DB 13082 GACATCATCACTCGGCTTCGATCGATCCCGAAGAG----- 13050
QY 253 LeuLeuLysLysIleCysAspAlaTyrTyrLeuProAlaTyrCysSerThrSer 270
DB 13049 -----TACGCGCGGCTTGGACGCGCAGCAGC 13023

RESULT 10
US-09-370-700-1/c
Sequence 1, Application US/09370700
Patent No. 6274350
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Turner, Jan R.
APPLICANT: Treadway, Patli J.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
FILE REFERENCE: 50489 DIV1
CURRENT APPLICATION NUMBER: US/09/370,700
EARLIER FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/36987
EARLIER FILING DATE: 1998-03-09
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 80161
TYPE: DNA
ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Alignment Scores:
Pred. No.: 2e-05 Length: 80161
Score: 152.50 Matches: 48
Percent Similarity: 38.38% Conservative: 28
Best Local Similarity: 24.24% Mismatches: 77
Query Match: 8.33% Indels: 45
Gaps: 4
DB: 4

US-09-857-613A-28 (1-350) x US-09-370-700-1 (1-80161)

QY 85 GlyAspHisMetHisGlyPheTyrAsp----- 94
DB 13517 GGGCGTCCCTCCACCGGCTACTGGGCGGGGTATCGGAGATCCCGTCCACA 13458
QY 95 -----SerAspSerThrValSerLeuSerAspHisArgAlaAlaGlnIleArgMetIle 112
DB 13457 CCGTGTCCGATGCTCCGCCACCACTGACCGAC-----CTGTTCACTGAC 13413
QY 113 GlnGluSerLeuArgPheAlaSerValSerGluGluArgSerIleTyrProIleSerIle 132
DB 13412 AAGGCCCGCGCTCCGCCGAGCG-----CACCTG 13383
QY 133 ValAspValGlyGlySerGlyIleGlyGlySerArgTyrLeuAlaLysLysPheGlyAla 152
DB 13382 TTCGACTGCGGCTGGCGCATGGGCGAGCCCGTAGTCCGTGGCGATGCCCGCGGCTT 13323

```

QY 153 ThrSerValGlyIleThrLeuSerProValGlnAlaGlnAlaArgAlaAsnAlaLeuAla 172
 Db 13322 CGAGTACCGGAAATCACCGTAAAGCCGAGATCTCGCCGCCGCCACAGGCTGCCCAAC 13263
 QY 173 AlaGlnGlyLeuAlaAspLysValSerPheGlnValAlaAlaAspAlaLeuGlnGlnProPhe 192
 Db 13262 GAGACCGGACGCGCGGAGCTTGTGATGTGATGTCAGACGCGGCCGACGCTGCCCTAC 13203
 QY 193 SerAspGlyGlnPheAspLeuValTrpSerMetGluSerGlyGlnHisMetProAspLys 212
 Db 13202 CCGGACGCTTCTTTCACGCGCGGAGTGCAGATCGCTCGTGCAGATCGTGACAGC 13143
 QY 213 AlaLysPheValGlyGlnLeuAlaArgValAlaAlaProGlyAlaIleIleIleVal 232
 Db 13142 GCCGCGCGGATCGCGGAGCTCCAGCAATCTCGGACCGCGCGGCTGCTCCGCA 13083
 QY 233 ThrTrpCysHisArgAspLeuGlyProAspGlnGlnSerLeuHisProTrpGlnLasp 252
 Db 13082 GACATCATCAGCTCGCGGTCGAGCTCCGCAAGG----- 13050
 QY 253 LeuLeuLysLysIleCysAspAlaIleTyrIleProAlaIleTrpCysSerThrSer 270
 Db 13049 -----TACGCGCGGCTTGGACGCGCAGCAGC 13023
 RESULT 11
 US-09-066-047-1
 Sequence 1, Application US/09066047A
 Patent No. 6306394
 GENERAL INFORMATION:
 APPLICANT: MURPHY, Cheryl
 BELTZ, Gerald A.
 STOREY, James
 TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND METHODS OF
 USE OF GRANDLOCYTIC ERHL1CH1A
 NUMBER OF SEQUENCES: 41
 CORRESPONDENCE ADDRESS:
 ADDRESS: HALE AND DORR LLP
 STREET: 60 State Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: United States
 ZIP: 02109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/066,047A
 FILING DATE: 24-Apr-1998
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/044,869
 FILING DATE: 25-Apr-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Superko, Colleen
 REGISTRATION NUMBER: 39,850
 REFERENCE/DOCKET NUMBER: 106,941.156
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 526-6000
 TELEFAX: (617) 526-5000
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 483 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:
 US-09-066-047-1

Alignment Scores:
 Pred. No.: 1,14e-05 Length: 4833
 Score: 138.00 Matches: 55
 Percent Similarity: 44.93% Conservative: 38
 Best Local Similarity: 26.57% Mismatches: 83
 Query Match: 7.54% Indels: 31
 DB: 4 Gaps: 8
 US-09-857-613a-28 (1-350) x US-09-066-047-1 (1-4833)
 QY 92 PheTyrAspSerAspSerThrValSerLeuSerAspHisArgAlaAlaGlnIleArgMet 111
 Db 3713 TCGTGAGATGGGAGAGCTTTTTCAGCTTTCACAGATGAATATCTTCCGCGTTACAGTAT 3772
 QY 112 IleGlnGluSerLeuArgPheAlaSerValSerGluGlnArgSerLysTrpProLysSer 131
 Db 3773 ATTCCTGAAATTTACAAAGGCTTACTACTACGTTAAAG----- 3814
 QY 132 IleValAspValGlyCysGlyIleGlyGlySerSerArgTrpLeuAlaLys--LysPhe 150
 Db 3815 CTTTGGATATGCGTTGTGTGTGTGTATTTGGCAAGCCATGGCAAGGCTTGTTT 3874
 QY 151 GlnAlaThrSerValGlyIleThrLeuSerProValGlnAlaGlnAlaArgAlaLeu 170
 Db 3875 ACTGTCACCTGAGATGATCCATGTAGAGAAAGAAATAGAGCTGCTAGACAGC----- 3928
 QY 171 AlaAlaAlaGlnGlyLeuAlaAspLysValSerPheGlnValAlaAspAlaLeuGln 190
 Db 3929 GCTGCTATGAAAGGCTTATATATAGAGTACCATTTACGATATAGAGTCTTTATATAC 3988
 QY 191 ProPheSerAspGlyGlnPheAspLeuValTrpSerMetGluSerGlyGlnHisMetPro 210
 Db 3989 TCCTCAGAGTGTCTTCTTACGATATCATCACCCTTATGAACTTGTAGACATATCCCT 4048
 QY 211 AspLysAlaIleSerPheValGlyGlnLeuAlaArgValAlaAlaProGlyAlaIleIle 230
 Db 4049 GATTTCAGTAATTTTACTTACTCTCTGTAAGTACTGAAACCTGAGGATGCTTTTC 4108
 QY 231 IleValThrTrpCysHisArgAspLeuGlyProAspGlnGlnSerLeuHisProTrpGlu 250
 Db 4109 ATTTCTACA-----CTAAACAGACTATCAATCCATG----- 4141
 QY 251 GlnAspLeuLeuLysLysIleCysAspAlaIleTyrIleProAlaIleTrpCysSerThr 270
 Db 4142 -----TTACTTGTAAGATAGCTCGCGAGTATATACTTC--GCATGGTCC----- 4184
 QY 270 rAspTyrValLysLeuLeuGlnSerLeuSerLeuGlnAspIleLysSerGluAspTrp 290
 Db 4185 -----CTCCTGGCAGCAGCAGCTGGAAGAAAGTTGTCAAGC-----CTTCA 4225
 QY 290 rArgPheValAlaProPhe 296
 Db 4226 GAGATTTCAGATGCCCTAT 4244
 RESULT 12
 US-08-961-527-222
 Sequence 222, Application US/08961527
 Patent No. 6420135
 GENERAL INFORMATION:
 APPLICANT: Charles Kunsch
 TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
 NUMBER OF SEQUENCES: 391
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
 COMPUTER: HP Vectra 486/33


```

Db 129 CACGACAGCTACACGTCGCGAGAACGTGCCCCGCTTTGGCCGCCGCTATTAATCAGC 198
QY 304 aPheThrTrpLysSclYLeuSerSerLeuLeuSerSerGlyGlnLysThrIleLysGlyAl 324
Db 189 GCTAAATGGAAGGAGGCTTACACCTCTGCTGTCGACGACCGGATGGAAGCATCAGATCAGAGCGC 248
QY 324 aLeuAlaMetProLeuMetIleGluGlyTyrLysLysAspLeuIleLysPheAlaIle 344
Db 249 GATGTGTATGCCGCTTAATGATCAGGCGCTACAGAAAGGCGCTCATCA-TTACACCATCAT 307
QY 344 eThr 345
Db 308 CACC 311

RESULT 3
US-09-938-842A-1815
: Sequence 1815, Application US/09938842A
: Patent No. US20020160378A1
: GENERAL INFORMATION:
: APPLICANT: Harper, Jeff
: APPLICANT: Krepis, Joel
: APPLICANT: Wang, Xun
: APPLICANT: Zhu, Tong
: TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
: TITLE OF INVENTION: SAME, AND METHODS OF USE
: FILE REFERENCE: SCRIPI300-3
: CURRENT APPLICATION NUMBER: US/09/938, 842A
: PRIOR FILING DATE: 2001-08-24
: PRIOR APPLICATION NUMBER: US 60/227, 866
: PRIOR FILING DATE: 2000-08-24
: PRIOR APPLICATION NUMBER: US 60/264, 647
: PRIOR FILING DATE: 2001-01-16
: PRIOR APPLICATION NUMBER: US 60/300, 111
: PRIOR FILING DATE: 2001-06-22
: NUMBER OF SEQ ID NOS: 5379
: SEQ ID NO 1815
: LENGTH: 1080
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana
US-09-938-842A-1815

Alignment Scores:
Pred. NO.: 8.89e-17 Length: 1080
Score: 226.00 Matches: 88
Percent Similarity: 42.21% Conservative: 61
Best Local Similarity: 24.93% Mismatches: 150
Query Match: 12.35% Indels: 54
DB: 9 Gaps: 13

US-09-857-613A-28 (1-350) x US-09-938-842A-1815 (1-1080)
QY 31 ValGlyProAlaGlySerTrpAlaProIleArgLysSerAlaLysSerSerGluArgGlyGlu 50
Db 70 CTAGGTCCAGCAGAACGCAAAAGGCAACGACCTCTGATCTCTCCGGC-----GGCTCA 123
QY 51 lIleValleuGlu-----GlnLysPro 57
Db 124 ATTCGCGGAGAAAAGTCAAGACACTATACCACTACTGCTCTTCTTCGCGAACA 183
QY 58 LysLysLys---AspAspLysLysLysLeuGlnLysGlyIleAlaGluPheTyrAspGluSer 76
Db 184 AAAAGAGATCGAATCAGCGCGAAGAAAGTACCTGCTCGCGACACGCTTCAACAACTTGTC 243
QY 77 SerGlyLeuTrpGluAsnIleTrpGlyAspHisMetHis-----Gly 91
Db 244 ACTGATATCTACGATGCGGATGGGATGGGACAAATCTTCCATTTCTCTCTCATATGCCCTGGA 303
QY 92 PheTyrAspSerAspSerThrValSerLeuSerAspHisArgAlaIleAlaGlnIleArgMet 111
Db 304 AAATCCGACAAAGAGCGCAC-----AGA 327
QY 112 lIleGlnGluSerLeuAlaGluPheAlaSerValSerGluGluArgSerLysTrpProLysSer 131

```

Db 328 ATCCAGAGAAATG-----GCCGTGATCTCATCAAACTGAACCGGACAAAG 378
 QY 132 ILevalAspValGlyCysGlyIleGlySerSerArgTrpLeuAlaLysLysPheGly 151
 Db 379 ATTCTTGACGCTGTGGCGGTGGGCGCATGAGACCATCGCGCCATTCCACG 428
 QY 152 AlaThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaAsnAlaLeuAla 171
 Db 439 GCCCAAGTCACTGATCACTATCAACAGATGACCAAGTGAACGACCAAGCTTCACAC 498
 QY 172 AlaAlaGlnGlyLeuAlaAspLysValSerPheGlnValAlaAspAlaLeuGlnPro 191
 Db 499 AAGAAAGCTGACCTGATTCTCTCTCAACGCTGTTGTGTAACCTTTTAAAGATGCCG 558
 QY 192 PheSerAspGlyGlnPheAspLeuValTrpSerMetGluSerGlyGlnHisMetProAsp 211
 Db 559 TTGATGATAAACACGTTTGCGAGCTTACTCATGACACTACGCTGCTCCCTCAAG 618
 QY 212 LysAlaLysPheValGlyGlnLeuAlaArgValAlaAlaProGlyAlaIleIleIle 231
 Db 619 CTGGAAGAAGTATACCTCGGAGATCTTCAGAGTATGAACCAAGATCTTGTCTGCTCC 678
 QY 232 ValThrTrpCysHisArgAspLeuGlyProAspGluGlnSerLeuHisProTrpGluGln 251
 Db 679 TACGAATGCGTCAACACTGAAAAATACAGACGATGACGACGACGACGACGACGACG 729
 QY 252 AspLeuLeuLysLysIleCysAspAlaValTrpLeuProAlaTrpCysSerThrSerAsp 271
 Db 730 GACGATATTAAGGATCGAGACAGACGACGACGACGACGACGACGACGACGACGACG 789
 QY 272 TyrValLysLeuLeuGlnSerLeuSerLeuGlnAspLysSerGluAspTrpSerArg 291
 Db 790 ATAGCCGTCGACGGCAAGAAAGTTGGGTGAGGTAGTGAAGGAAGAAAGTTGGCTAAA 849
 QY 292 PheValAlaLysProPheTrpProAlaValIle-----ArgSerAlaPheThrTrpLys 308
 Db 850 CCACCCCTTAACCGCTGCGTGAACCGCTTAAGATGGAGATGGCTTAT---TGAGGA 906
 QY 309 Gly-----LeuSerSerLeuLeuSerSerGlyGlnLysThrIleLysGlyAlaLeuAla 326
 Db 907 AACCATGTTGCTGCTGATCTTCTTCTGCTATTGGGCTCTCTCAAGAAAGACGTTGAT 966
 QY 327 MetProLeuMetIle-----GluGlyTrpLysLysAspLeu 338
 Db 967 GTTCATAGATGTTGTTAAGCTGCTGATATTGACACGAGAGTGGTGAAGCTGATC 1026
 QY 339 IleLys---PheAlaIleIleThrCysArgLysProGlu 350
 Db 1027 TTCTCTCGATGATGATTCTCTGTAGAAACACGAG 1065

RESULT 4

US-09-779-144A-6
 : Sequence 6, Application US/09779144A
 : Patent No. US20020148006A1
 : GENERAL INFORMATION:
 : APPLICANT: Nes, David W.
 : TITLE OF INVENTION: Transgenic Plants with Modified Sterol Compositions
 : FILE REFERENCE: 11899.0198.DVDS01 MOBT:198
 : CURRENT APPLICATION NUMBER: US/09/779,144A
 : PRIOR APPLICATION NUMBER: US 09/106,926
 : PRIOR FILING DATE: 1998-06-29
 : NUMBER OF SEQ ID NOS: 16
 : SOFTWARE: PatentIn version 3.0
 : SEQ ID NO 6
 : LENGTH: 1497
 : TYPE: DNA
 : ORGANISM: Zea mays
 US-09-779-144A-6

Alignment Scores: 1,47e-16 Length: 1497
 Pred. No.: 226.00 Matches: 81
 Score:

Percent Similarity: 38.77% Conservative: 45
 Best Local Similarity: 24.92% Mismatches: 121
 Query Match: 12.35% Indels: 78
 DB: 10 caps: 10

US-09-857-613a-28 (1-350) x US-09-779-144A-6 (1-1497)

QY 69 ILeAlaLubPheTyrAspGluSerSerGlyLeuTrpGluAsnIleTrpGlyAspHisMet 88
 Db 216 GTTAAATAATACATATGATCTTCCACATAGCTTCTATGATAGTATGGTGGGATTCCTTC 275
 QY 89 HisHisGlyPheTyrAspSerAspSerThrValSerLeuSerAspHisArgAlaAlaGln 108
 Db 276 CACTTCTCTCAACATGAAATGAGAAATCCCTTACGTGAAGCATCAACGACGACATGACAT 335
 QY 109 IleArgMetIleGlnGlnSerLeuArgPheAlaSerValSerGluGlnArgSerLysTrp 128
 Db 336 TTCTCTCCCTGCAACTGCTTGTG-----AACCA 365
 QY 129 ProLysSerIleValAspValGlyCysGlyIleGlySerSerArgTrpLeuAlaLys 148
 Db 366 GGAATGAAGGTTTATGATGTGGCTGTGATAGTGTGACCATGACGACGACGACGACGACG 422
 QY 149 LysPheGlyAlaThrSerVal---GlyIleThrLeuSerProValGlnAlaGlnArgAla 167
 Db 423 AGATTAGCTCAACTTCAGTACCGGATTTGAATAAACGAATACCAAGATTAACGACGAGGA 482
 QY 168 AsnAlaLeuAlaAlaGlnGlyLeuAlaAspLysValSerPheGlnValAlaAspAla 187
 Db 483 AAGAGCTCAACCGTTTACGCAATTAATGCAACATGATGATTTGTCAACGCGGACTTC 542
 QY 188 LeuGlnGlnProPheSerAspGlyGlnPheAspLeuValTrpSerMetGluSerGlyGlu 207
 Db 543 ATGAAGATGCCGTTGCGATGACAAACACTTTATGCTGTTTACGCCATGACGACCAACATGT 602
 QY 208 HisMetProAspLysAlaLysPheValGlyGlnLeuAlaArgValAlaAlaProGlyAla 227
 Db 603 CATGACCTGATCCAGTTGATGCTTACACGAGATATATGCTGTTGGAACCTGCGCAG 662
 QY 228 IleIleIleIleValIleThrCys-----His 236
 Db 663 TCGTTTCCCGTACAGAGTGGTGTGTTAGCATCATGATCTTCAACATGCAACCCAG 722
 QY 237 Arg-----AspLeuGly-----ProAspGluGlnSerLeuHis 247
 Db 723 AAAAGATCAAGATGAATATGACCTTGCAATGGCCCTGCAGATATCAGAAGCACTGG 782
 QY 248 Pro-----TrpGluGlnAspLeu 254
 Db 783 CAATGTCGCCGCGCAGTAAAGACCGCGGTTTGAGTTGTTGGATAGATGATCTTCT 842
 QY 255 LysLysIleCysAspAlaValTrpLeuProAlaTrpCysSer----- 268
 Db 843 GAAATTTCTCCCTTGCTTGTGCTTGTGCTTGTGATCCATGCAACCGCAATCTCCCTGATAGC 902
 QY 269 -----ThrSerAspGlyTrpValLysLeuLeuGlnSer 278
 Db 903 TTCCGTTGACCTGTGCGACGATGATTAACCCGACCAATGATCAAGCCCTGGAGTAC 962
 QY 279 LeuSerLeuGlnAspIleLysSerGluLysProPheArgPheValAlaProPheTrpPro 298
 Db 963 GTTGCTGTGCTCGCAGAGGAGTGAGGCTCTACTTTCCTGAG----- 1010
 QY 299 AlaValIleArgSerAlaPheThrTrpLysGlyLeuSerSerLeuLeuSerSerGlyGln 318
 Db 1011 -----AAGCTGCAAGAGGCTGCTGAGCGCGCAAG 1043
 QY 319 LysThrIleLysGlyAlaLeuAlaMetProLeuMetIleGluGlyTrpLysLysAspLeu 338
 Db 1044 AAGGAGATC-----TTCAAGCCCAATGATCTTTTGTGTTGGAAGCCTCTT 1091
 QY 339 IleLysPheAlaIle 343
 : : : : :

Db 1092 CTGGAATGACCTCTT 1106

RESULT 5

US-09-938-842A-1856

Sequence 1856, Application US/0938842A

Patent No. US20020160378A1

GENERAL INFORMATION:

APPLICANT: Harper, Jeff

APPLICANT: Krepes, Joel

APPLICANT: Wang, Xun

APPLICANT: Zhu, Tong

TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

TITLE OF INVENTION: SAME, AND METHODS OF USE

FILE REFERENCE: SCRIPI300-3

CURRENT APPLICATION NUMBER: US/09/938,842A

CURRENT FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/227,866

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: US 60/264,647

PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/300,111

PRIOR FILING DATE: 2001-06-22

NUMBER OF SEQ ID NOS: 5379

SEQ ID NO 1856

LENGTH: 1086

TYPE: DNA

ORGANISM: Arabidopsis thaliana

US-09-938-842A-1856

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-857-613a-28 (1-350) x US-09-938-842A-1856 (1-1086)	3,35e-15	212.50	42.54%	25.40%	1086	80	54	144	37	11

Db 55 GlnLysProLysLys---AspAspLysLysLysLeuGlnLysGlyIleAlaGluPheTyr 73

Db 175 CGCGCGTCAAAAGAAATCGAAGACCGCGAAGAGTTCCACATTCGTCGACACATTTTAC 234

QY 74 AspGluSerSerGlyLeuTrpGluAsnIleTrpGlyAspHisMetHisGlyPheTyr 93

Db 235 AATCTCGTCACCGACATATTCAGATGGGATGGGACATCTCCAC-----TTCTCA 288

QY 94 AspSerAspSerThrValSerLeuSerAsp-----HisArgAlaAlaGlnIle 109

Db 289 CCATCATATCCCGGAAATCTCACAAGACGCCGCCCTCCACGAGATGCCGTA 348

QY 110 ArgMetIleGlnLysLeuSerLeuArgPheAlaSerValSerGluGluArgSerLysTrpPro 129

Db 349 GATCTGATCCAA-----GTCAACCTGGT 372

QY 130 LysSerIleValAspValGlyCysGlyIleGlyLysSerArgTyrIleAlaLysLys 149

Db 373 CAAAGATCTGATGACGCGATCGGTCGTCGCGGTCGATCGACGATTCATCTCAC 432

QY 150 PheGlyAlaThrSerValGlyIleThrLeuSerProValGlnAlaGlnAlaAsnAla 169

Db 433 TCCGAGCTACCTAGCTGCGGATTCACATTAACGATGATGAACAGCTCGCTC 492

QY 170 LeuAlaAlaIleGlnLysLeuAlaAspLysValSerPheGlnValAlaAlaAspAlaLeuGln 189

Db 493 CACAATTAAGAAAGCTGCTCGACGCGCTTTCGAGGTCTGTGTAACTTCCTCAG 552

QY 190 GlnProPheSerAspGlyGlnPheAspLeuValTrpSerMetGluSerGlyGlnHisMet 209

Db 553 ATGCGCTGATGACAAAGACTTTCGACGCGGTATTCATTCGACGACGCTGACGCG 612

QY 210 ProAspLysAlaLysPheValGlyGlnLeuAlaArgValAlaAlaLapProGlyAlaIleTle 229

Db 613 CCGAAGCTGAGAGAGTATCGACAGATCTACAGGCTGTTGAACCCGGATCTATGTAT 672

QY 230 IleIleValThrTrpCysHisArgAspLeuGlyProAspGluGlnSerLeuHisProTrp 249

Db 673 GTGTCTGACAGATGCGTTCAGACGAGCAAAATTTAAGCGGAGATGACGAACAC----- 726

QY 250 GluGlnAspLeuLeuLysLysIleCysAspAlaTrpTyrTrpLeuProAlaTrpCysSerThr 269

Db 727 ---GTGAGCTTAATCCAAAGGATTTGAGAGAGCGCATCGTTACACGCGCTTAGCGCTTAC 783

QY 270 SerAspPyrValLysLeuLeuGlnSerLeuSerLeuGlnAspIleLysSerGluAspTrp 289

Db 784 GTGATATTAATCTGAGACGCTAAAGAGTTGGCTTTGATGATGAGAGAGAGACATCTG 843

QY 290 SerArgPheValAla---ProPheTrpProAlaValIle-----ArgSerAlaPheThr 306

Db 844 GCGAGTCCACCGCGCTGAGCGCGTGTGACACGCTTAAGATGGGTAGCGCTTCTAT--- 900

QY 307 TrpLysGly-----LeuSerSerLeuLeuSerSerGlyGlnLysThrIleLysGlyAla 324

Db 901 TGGAGCAATCACAATTGCTGCTCAGATTGTCAGCGCTTGAGTCTCTCTTAAGCAACT 960

QY 325 LeuAlaMetProLeuMetIle-----GluGlyTyrLysLys 336

Db 961 GTTGATGTTTCATGAGATGTTGTTTAAAGCTGCTGATTTATTTGACCAAGAGAGTGAAACC 1020

QY 337 AspLeuIleLys---PheAlaIleIleThrCysArgLysProGlu 350

Db 1021 GAATATTTCTCTCCGATCATATGATCTCTCGCAAAACCGGAG 1065

RESULT 6

US-09-801-368-95

Sequence 95, Application US/09801368

Patent No. US20020128250A1

GENERAL INFORMATION:

APPLICANT: Busby, Robert

APPLICANT: Call, Brian

APPLICANT: Hecht, Peter

APPLICANT: Holzman, Doug

APPLICANT: Madden, Kevin

APPLICANT: Maxon, Mary

APPLICANT: Milne, Todd

APPLICANT: No. US20020128250A1man, Thea

APPLICANT: Royer, John

APPLICANT: Salama, Sofie

APPLICANT: Sherman, Amir

APPLICANT: Silva, Jeff

APPLICANT: Summers, Eric

TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fu

FILE REFERENCE: 109272.147

CURRENT APPLICATION NUMBER: US/09/801,368

CURRENT FILING DATE: 2001-03-07

PRIOR APPLICATION NUMBER: US 09/487,558

PRIOR FILING DATE: 2000-01-19

PRIOR APPLICATION NUMBER: US 60/160,587

NUMBER OF SEQ ID NOS: 440

SOFTWARE: PatentIn version 3.0

SEQ ID NO 95

LENGTH: 1152

TYPE: DNA

ORGANISM: Saccharomyces cerevisiae

US-09-801-368-95

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-857-613a-28 (1-350) x US-09-801-368-95 (1-1152)	4.09e-14	203.50	37.99%	21.79%	1152	78	58	149	73	12

```

QY 42 SerAlaIaSerSerGluArgGlyGluIleValLeuGluGlnLysProLysLysAsp 61
DB 106 AACCTCTCCCAAGAGAGCCGCTTCAAGAACTACTTGAACAAATTGGATGTAGAACCAT 165
QY 62 LysLysLysLeuGlnLysGlyIleAlaGluPheTyrAspLysSerSerGlyLeuTyrGlu 81
DB 166 AAGATGCCGAAGAACGTCGCTT---GAGCATTTATATGACGCCACATCTCTACTAT 222
QY 82 AsnIle-----TyrGlyAspHisMetHisGly---Phe 92
DB 223 AACGTCCTTACAGATTCTATGATATGATGTTGGGTCCTCTTCCATTTCCAGCAATTT 282
QY 93 TyrAspSerSer-----ThrValSerLeuSerSerHisAlaAlaGluIle 109
DB 283 TATTAAGTCGAGTTTCTGCTGCTCCATGACAGACATGACATTTATTAGCTTACAG 342
QY 110 ArgMetIleGlnLysLeuArgPheAlaSerValSerGluLysSerLysTyrPro 129
DB 343 GCTGCTATTCAAGAGGC----- 360
QY 130 LysSerIleValAspValGlyGlyIleGlyLysSerSerArgTyrLeuAlaLysLys 149
DB 361 GATTAGTTCTCGACGTTGTTGTTGGGGCCCAAGACAGATTCAGATTT 420
QY 150 PheGlyAlaThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaAsnAla 169
DB 421 ACCGCTTGAACGTCATCGCTTAAACATTAACGATTACCAATTGCGCAAGCCAAATAT 480
QY 170 LeuAlaIaIaIaGlnLysLeuAlaAspLysValSerPheGlnAlaIaAspAlaLeuGln 189
DB 481 TAGCGTAAATAATACATTTGACTGACCAATGACTTTGTAAGGCTTTTCATGAAA 540
QY 190 GlnProPheSerAspLysGlnPheAspLeuValTyrSerMetGluSerGlyGlnHisMet 209
DB 541 ATGCAATTCGAGAAACACTTTCGACAAAGTTTATGCAATTCAGGCCACATCTCACGCT 600
QY 210 ProAspLysAlaLysPheValGlyGluLeuAlaArgValAlaAlaProGlyAlaIleIle 229
DB 601 CCAAAATTGAAGGCTGATACAGCCGAAATCTACAGGTTTGAACGGGTGTACCTTT 660
QY 230 IleIleValThrTyrPysHisArgAspLeuGlyProAspGluGlnSerLeuHisProTyr 249
DB 661 GCGTTTACGATGGGTATGACTGATTAATATGACGAAACATCTCTCAACAT----- 714
QY 250 GluGlnAspLeuLeuLysLysIleCysAspAlaTyrTyr-----Leu 263
DB 715 -----AGAAAGATC-----GCTTATGAATTTGAAGTACGTGATGCTATC 753
QY 264 ProAlaTyrPysSerThrSerAspTyrValLysLeuGlnSerLeuSerLeuGlnAsp 283
DB 754 CCAAGATGTTCCATGTCGACGCTGCTAGAAACATGAAAGCTGCTTTCGAAAGTC 813
QY 284 IleLysSerGluAsp-----TyrSerArgPheValAla 294
DB 814 CTCGTTACGAAAGACTCGCGGACATGATGAAATCCCTTGATTTACCCATTAAT 873
QY 295 ProPheTyrProAlaVal-----IleArgSerAlaPhe 305
DB 874 GGTGAGTGAAGTACGTTCAAAACTTAGCTATTTGCCACATTTTTCGAACTCTTAC 933
QY 306 ThrTyrPysGlyLeuSerSer-----LeuLeuSer 315
DB 934 TTGGGTACACATTTTACACAGCATGCTTACTGTAATGAGAAATTAAGTCTACGCCCA 993
QY 316 SerGlyGlnLysThrIleLysGlyAlaLeu-----AlaMetProLeuMetIleGlu 332
DB 994 GAAGTTCACAGAGATTCAGTCTGCTCTAGAAATGCTCGGCTTGTTAGTCTCCCGGT 1053
QY 333 GlyTyrLysLysAspLeuIleLysPheAlaIleIleThrCysArgLysProGlu 350
DB 1054 GGTAGTCCAAAGTTATTCATCTCAATGATGCTTTTCGTCGTAGAACCCAGAA 1107

```

```

RESULT 7
US-09-779-144A-1
: Sequence 1, Application US/09779144A
: Patent No. US20020148006A1
: GENERAL INFORMATION:
: APPLICANT: Nes, David W.
: TITLE OF INVENTION: Transgenic Plants with Modified Sterol Compositions
: FILE REFERENCE: 11899.0198, DIV501, MOBT:198
: CURRENT APPLICATION NUMBER: US/09/779,144A
: PRIOR APPLICATION NUMBER: 2001-02-08
: PRIOR FILING DATE: 1998-06-29
: NUMBER OF SEQ ID NOS: 16
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 1
: LENGTH: 1320
: TYPE: DNA
: ORGANISM: Saccharomyces cerevisiae
US-09-779-144A-1

```

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
5,05e-14	203.50	37.99%	21.79%	11.12%	1320	78	58	149	73	12

US-09-857-613a-28 (1-350) x US-09-779-144A-1 (1-1320)

```

QY 42 SerAlaIaSerSerGluArgGlyGluIleValLeuGluGlnLysProLysLysAsp 61
DB 177 AACCTCTCCCAAGAGAGCCGCTTCAAGAACTACTTGAACAAATTGGATGTAGAACCAT 226
QY 62 LysLysLysLeuGlnLysGlyIleAlaGluPheTyrAspLysSerSerGlyLeuTyrGlu 81
DB 237 AAGATGCCGAAGAACGTCGCTT---GAGCATTTATATGACGCCACATCTCTACTAT 293
QY 82 AsnIle-----TyrGlyAspHisMetHisGly---Phe 92
DB 294 AACGTCCTTACAGATTCTATGATATGCTTGGGTCCTCTTCCATTTCCAGCAATTT 353
QY 93 TyrAspSerSer-----ThrValSerLeuSerSerHisAlaAlaGluIle 109
DB 354 TATTAAGTCGAGAGTTTCCGCTGCTGATGACAGACATGACATTTATTAGCTTACAG 413
QY 110 ArgMetIleGlnLysLeuArgPheAlaSerValSerGluLysSerLysTyrPro 129
DB 414 GCTGCTATTCAAGAGGC----- 431
QY 130 LysSerIleValAspValGlyGlyIleGlyLysSerSerArgTyrLeuAlaLysLys 149
DB 432 GATTAGTTCTCGACGTTGTTGTTGGGGCCCAAGACAGATTCGAAATTT 491
QY 150 PheGlyAlaThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaAsnAla 169
DB 492 ACCGCTTGAACGTCATCGCTTAAACATTAACGATTACCAATTTGCCAAGGCAATAT 551
QY 170 LeuAlaIaIaIaGlnLysLeuAlaAspLysValSerPheGlnValAlaIaAspAlaLeuGln 189
DB 552 TAGCGTAAATAATACATTTAGTGTGACCAATGACATTTGTAAGGCTATTTTCGAA 611
QY 190 GlnProPheSerAspLysGlnPheAspLeuValTyrSerMetGluSerGlyGlnHisMet 209
DB 612 ATGATTTTCGAGAAACACTTTCGACAAAGTTTATGCAATTTGAGGCCACATGTCACGCT 671
QY 210 ProAspLysAlaLysPheValGlyGluLeuAlaArgValAlaAlaProGlyAlaIleIle 229
DB 672 CCAAAATTGAAGGCTGATACAGCAATCTTCAAGAGTTTGAACGGGTGTACCTTT 731
QY 230 IleIleValThrTyrPysHisArgAspLeuGlyProAspGluGlnSerLeuHisProTyr 249
DB 732 GCTGTTACGAATGGGTATGATGATGATTAATATATGACGAAACATCTCTCAACAT----- 785

```

```

OY 250 GluInAspLeuLysLysIleCysAspAlaTyrTyr-----Leu 263
Db 786 -----AGAAAGATC-----GCTTATGAAATGAACTAGTGATGATTC 824
OY 264 ProAlaTrpCysSerThrSerAspTyrValLysLeuGlnSerLeuSerLeuGlnAsp 283
Db 825 CCAAGATGTTCCATGTCGACGTGCTAGGAAAGCATTGAGAACTGCTTTCACATC 884
OY 284 IleLysSerGluAsp-----TrpSerArgPheValAla 294
Db 885 CTCGTTACGACGACCTGCGGACCAATGATGAAATCCCTTGATTAACCATTAAT 944
OY 295 ProPheTrpProAlaVal-----IleArgSerAlaPhe 305
Db 945 GGTGACGTGAAGTACCTTCAAAACTAGCTAATTTGGCCACATTTTCAGAACTCTTAC 1004
OY 306 ThrTrpLysGlyLeuSerSer-----LeuLeuSer 315
Db 1005 TTGGGTAGACAAATTTACTACAGCAATGTTACTGTAATGAAAAATTAAGTCTACGCCCA 1064
OY 316 SerGlyGlnLysThrIleLysGlyAlaLeu-----AlaMetProLeuMetIleGlu 332
Db 1065 GAAGGTTCCAGAGAGTACTGCTCTAGAAAATGCTGCGGTTGTTAGTTGCCGCT 1124
OY 333 GlyTyrLysLysAspLeuIleLysPheAlaIleIleThrCysArgLysProGlu 350
Db 1125 GGTAACTCCAGATTATTCACCTCAATGATGCTTTTCCTGCTAGGAGCCAGAA 1178

RESULT 8
US-09-779-144A-5
: Sequence 5, Application US/09779144A
: Patent No. US20020148006A1
: GENERAL INFORMATION:
: APPLICANT: Nes, David W.
: TITLE OF INVENTION: Transgenic Plants with Modified Sterol Compositions
: FILE REFERENCE: 11899, 0198, DVUS01 MORT:198
: CURRENT APPLICATION NUMBER: US/09/779,144A
: CURRENT FILING DATE: 2001-02-08
: PRIOR APPLICATION NUMBER: US 09/106,926
: PRIOR FILING DATE: 1998-06-29
: NUMBER OF SEQ ID NOS: 16
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 5
: LENGTH: 1320
: TYPE: DNA
: ORGANISM: Saccharomyces cerevisiae
US-09-779-144A-5

Alignment Scores:
Pred. No.: 5, 05e-14 Length: 1320
Score: 203.50 Matches: 78
Percent Similarity: 37.99% Conservative: 58
Best Local Similarity: 21.79% Mismatches: 149
Query Match: 11.12% Indels: 73
DB: 10 Gaps: 12

US-09-857-613a-28 (1-350) x US-09-779-144A-5 (1-1320)
OY 42 SerAlaAlaSerSerGluArgGlyGluIleValLeuGlnLysProLysLysAspAsp 61
Db 177 AACTGTGCCAAAAGAGAGCGCTTCAGAAAGTACGTAAGAAATTTGGATGTRGAAACCGAT 236
OY 62 LysLysLysLeuGlnLysGlyIleAlaGluPheTyrAspGluSerSerGlyLeuTrpGlu 81
Db 237 AAAGATGCCAGAGAACGTCGCTT--GAGGATTATTAAGAACCCACACATTCCTACTAT 293
OY 82 AsnIle-----TrpGlyAspHisMetHisIleGly---Phe 92
Db 294 AACGTGTTACAGATTCTTATGATATGTTGGGTTCTCTTCCATTTCACACAGATT 353
OY 93 TyrAspSerAspSer-----ThrValSerLeuSerAspHisArgAlaIleGlnIle 109

```

```

Db 354 TATAAGGTGAGACTTTCCTGCTCCCTCAGTACGAAACATGACATTAATTAGCTTACAG 413
OY 110 ArgMetIleGlnGlnSerLeuArgPheAlaSerValSerGluGlnArgSerLysTrpPro 129
Db 414 GCTGTATTCAAAAGAGC-----
OY 130 LysSerIleValAspValAlaGlyCysGlyIleGlyLysSerSerArgTyrLeuAlaLysLys 149
Db 432 GATTAGATCTTCACGCTTGTTGCTGTGCTGGGCGCCAGCAAGACAGATTGACAGATT 491
OY 150 PheGlyAlaThrSerValGlyIleThrLeuSerProValGlnAlaIleArgAlaAsnAla 169
Db 492 ACCGTTGTAACGTCATCGCTCAACATTAACATTAACATTAACATTAACATTAACATTA 551
OY 170 LeuAlaAlaIleGlnGlyLeuAlaLysValSerPheGlnValAlaLysAlaLeuGln 189
Db 552 TACGCTAAAAAATACAAATTGAGTGACCAATGCAATGTTGAAGGCGATTTACAGAA 611
OY 190 GlnProPheSerAspGlyGlnPheAspLeuValTrpSerMetGlnSerGlyGlnHisMet 209
Db 612 ATGCAATTTCCAGAAAACACTTTCGACAAAGTTTATGCAATTTGAGGCCACATCTCACGCT 671
OY 210 ProAspLysAlaLysPheValGlyGlnLeuAlaArgValAlaIleAlaIleIle 229
Db 672 CCAAAATTAGAGGTGATACAGCGAATCTACAGGTTTGAACCGGCTGACCTT 731
OY 230 IleIleValThrTrpCysHisArgAspLeuGlyProAspGlnGlnSerLeuHisProTrp 249
Db 732 GCTGTTACGAATGCTGTAATGCTGATTAATATGACGAAACATCTCGAACAT----- 785
OY 250 GluInAspLeuLysLysIleCysAspAlaTyrTyr-----Leu 263
Db 786 -----AGAAAGATC-----GCTTATGAAATGAACTAGTGATGATTC 824
OY 264 ProAlaTrpCysSerThrSerAspTyrValLysLeuGlnSerLeuGlnAsp 283
Db 825 CCAAGATGTTCCATGTCGACGTGCTAGGAAAGCATTGAGAACTGCTTTCACATC 884
OY 284 IleLysSerGluAsp-----TrpSerArgPheValAla 294
Db 885 CTCGTTACGACGACCTGCGGACCAATGATGAAATCCCTTGATTAACCATTAAT 944
OY 295 ProPheTrpProAlaVal-----IleArgSerAlaPhe 305
Db 945 GGTGACGTGAAGTACCTTCAAAACTAGCTAATTTGGCCACATTTTCAGAACTCTTAC 1004
OY 306 ThrTrpLysGlyLeuSerSer-----LeuLeuSer 315
Db 1005 TTGGGTAGACAAATTTACTACAGCAATGTTACTGTAATGAAAAATTAAGTCTACGCCCA 1064
OY 316 SerGlyGlnLysThrIleLysGlyAlaLeu-----AlaMetProLeuMetIleGlu 332
Db 1065 GAAGGTTCCAGAGAGTACTGCTCTAGAAAATGCTGCGGTTGTTAGTTGCCGCT 1124
OY 333 GlyTyrLysLysAspLeuIleLysPheAlaIleIleThrCysArgLysProGlu 350
Db 1125 GGTAACTCCAGATTATTCACCTCAATGATGCTTTTCCTGCTAGGAGCCAGAA 1178

RESULT 9
US-09-779-144A-3
: Sequence 3, Application US/09779144A
: Patent No. US20020148006A1
: GENERAL INFORMATION:
: APPLICANT: Nes, David W.
: TITLE OF INVENTION: Transgenic Plants with Modified Sterol Compositions
: FILE REFERENCE: 11899, 0198, DVUS01 MORT:198
: CURRENT APPLICATION NUMBER: US/09/779,144A
: CURRENT FILING DATE: 2001-02-08
: PRIOR APPLICATION NUMBER: US 09/106,926
: PRIOR FILING DATE: 1998-06-29
: NUMBER OF SEQ ID NOS: 16
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 3

```



```

OY      184 ValAlaSpALalLeuGInGLnPrPheserAspLyGIlnPheaspleValITrpSerMet 203
          ||||| : : : : : ||||| : : : : :
Db      163 TGTGGCGATTCCACTAAAAAACATACCTAGAAATCATTTGGTGAATTCATATTCGCCGT 222
OY      204 GlusecrglyluHismetProaspIySalyspheValIGlyGUleuAlaArgValAla 223
          : : : : : |||||
Db      223 GACACATGTTTAACAATCAATCAAGAATAAACCACCACTATTCAGATCATTTTACAAATGGTTG 282
OY      224 AlaProglYAlaIlellelleValIThrTrpCysHisArgspLeuGlyPro 241
          ||| : : : ||||| : : : |||
Db      283 AAGCGTGAGGTACACTTATTATTACTGATTTCGAAAAGTAGAAGAGTCCA 336

RESULT 12
US-10-260-877-53
; Sequence 53, Application US/10260877
; Publication No. US20030021813A1
GENERAL INFORMATION:
APPLICANT: Abbott Laboratories
APPLICANT: Chovan, Linda E.
APPLICANT: Hessler, Paul E.
APPLICANT: Reich, Karl A.
TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME
TITLE OF INVENTION: SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF
TITLE OF INVENTION: 'ESSENTIAL' GENES,
FILE REFERENCE: 6565 US P1
CURRENT APPLICATION NUMBER: US/10/260,877
CURRENT FILING DATE: 2002-09-30
PRIOR APPLICATION NUMBER: US/09/649,145
PRIOR FILING DATE: 2000-08-25
NUMBER OF SEQ. ID NOS: 137
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 53
LENGTH: 753
TYPE: DNA
ORGANISM: H. influenzae
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(753)
OTHER INFORMATION: HI-0095
US-10-260-877-53

Alignment Scores:
Pred. No.: 4,02e-07 Length: 753
Score: 141.00 Matches: 47
Percent Similarity: 44.68% Conservative: 37
Best Local Similarity: 25.00% Mismatches: 74
Query Match: 7,70% Indels: 30
DB: 9 Gaps: 5

US-09-857-613A-28 (1-350) x US-10-260-877-53 (1-753)
OY      130 LysSerLleValaspValIGlyCySGlylleglylYserSerArgrlyrLeuAlalySlys 149
          ||| : : : : : ||| : : : : :
Db      118 AAAAAGATGTGGAGGTGCCTGCTGAATATGGGAGCAGCATCAATGTGATTGGCAAACAA 177
OY      150 PheGlylaThrSerValIGlylIethrLeuSerProvalGINalagInatgAlaAsnaLa 169
          ||||| : : : : : |||
Db      178 TTTCGTGTCTATTTGAAGGCTGTGATTTAATCAATAAATGCCCTTACGAAAAGCACAGCA 237
OY      170 LeuAlaAlaInglyLeuAlaAsplysValSerPheGInAlaAlaaspAlaLeuGln 189
          ||| : : : : : |||
Db      228 AATATTGAAGCAAAATGGCTTCAGGAATAAATTCATGTACAGCGTGCAGCAATGCATGAMG 297
OY      190 GlnProPheSerAspGlyGlnPheaspLeuValITrpSerMetGluserglyluHismet 209
          ||||| ||| : : : : : |||
Db      298 TTGCCTTTCGAGATGAAGATTTCATATTTGCATCATCAATGAAGCGATGCTCAATGTTA 357
OY      210 Pro-----AsplysAlaLysPheValIGlyGUleuAlaArgValAlaAlaProGlyAla 227
          ||| ||| ||| : : : ||| ||||| |||
Db      358 CCCGTGGAAGCAAGAAAGCAAAAGCCATTTGCAGATATTTTTCGAGCTGTAAACCCCAATGCT 417
OY      228 IleIleIleIleValIThrTrpCysHisArgspLeuGlyProAspGluIndSerLeuHis 247
          : : : : : ||| ||| : : : : :

```

```

Db 418 TTATTGCTT-----ACTCAGCATGTTATGCTGCGGAATGATCATCAACTATT--- 468
Qy 248 ProtrpGlulnAspLeuLysLysLleCysAspAlaTyrTyrLeuProAlaIrrpys 267
Db 469 -----CTAGAAATATATGCCAAGCGAT----- 492
Qy 268 SerThrSerAspTyrValLysLeuLeuGlnSerLeuSerLeuGlnAspLleLysSerGlu 287
Db 493 -----AACGTGACTGTCACGCCCATTAACCAAGAT 522
Qy 288 AspTrpSerArgPheValAlaProPheTrpProAlaValIleArgSerAlaIlePheThrTrp 307
Db 523 GGATGGAAAGCGATA-----TTCCAAGAAAGTGTTTAGAAATGTTGATACCTTTC 573
Qy 308 LysGlyLeuSerSerLeuLeuSer 315
Db 574 TCTGTGAGATGACATTACTTTC 597

RESULT 13
US-09-845-248-1
: Sequence 1, Application US/09845248
: Patent No. US2002042106A1
: GENERAL INFORMATION:
: APPLICANT: Nampoothiri, Madhavan
: TITLE OF INVENTION: Nucleotide Sequences Which Code for the CMA Gene
: FILE REFERENCE: 032301 WD 1161
: CURRENT APPLICATION NUMBER: US/09/845,248
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 1
: LENGTH: 1801
: TYPE: DNA
: ORGANISM: Corynebacterium glutamicum
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (248)..(1600)
: OTHER INFORMATION:
US-09-845-248-1

Alignment Scores:
Pred. No.: 2,02e-06 Length: 1801
Score: 140.00 Matches: 83
Percent Similarity: 36.06% Conservative: 45
Best Local Similarity: 23.38% Mismatches: 120
Query Match: 7.65% Indels: 107
Db: 10 Gaps: 15

US-09-857-613A-28 (1-350) x US-09-845-248-1 (1-1801)

Qy 53 LeuGlnGlnLysPro-----LysLysAspAsp 61
Db 674 ATGGAAACAGCCCTGCATGGCGTAAAGCATCATCAACGCCGTACAGCCACTCG 733
Qy 62 LysLysLysLeuGlnLysGlyIleAlaGluPheTyrAspLysSerGlyLeuTrpGlu 81
Db 734 AATATCCCGGACAGAAAGCAACCATTAACCTACAGAGAGCGGAGATGTTCTTCTCC 793
Qy 82 AsnIleTrpGlyAspHisMetHis-----GlyPheTyrAspSerAspSerThrVal 99
Db 794 CTGTTTATAGATGATTCCTACCTATACCTCGCGCTATTATCAACGCAGAAATCA--- 850
Qy 100 SerLeuSerAspHisArgAlaIleAlaGlnIleArgMetIleGlnLysSerLeuArgPheAla 119
Db 851 AGTTTGAAGAAAGCCCAAGAAACAAATACCGCTCATCTTTGAAAAACTGGCTG--- 907
Qy 120 SerValSerGluGlnArgSerIleValAspValGlyCysGlyIle 139
Db 908 -----AAAGAAAGCGATCCCTCTAGACGTGGATCGCGTTGG 946
Qy 140 GlyLysSerArgTyrTyrLeuAlaLysPheGlyAlaThrSerValGlyIleThrLeu 159
Db 947 GGAGCGCATGTCCTTAC---GCCGCCAAACACGCTGTGAAGAACCATCGAGTTACGCTG 1003

```

```

Qy 160 SerProValGlnAlaGlnArgAlaAsnAlaLeuAlaIleAlaGlnGlyLeuAlaAspLys 179
Db 1004 TCTGAACAGCAATATGATGTGGGTCAACAGATCAACACCCAAAGTTTGAAGACCTC 1063
Qy 180 ValSerPheGlnValAlaAspAlaLeuGlnIlePheSerAspGlyGlnPheAspLeu 199
Db 1064 GCGAAATTCCTTCATGATGATTACCGCATTTCCA-----GAACCTGATTCGATGCG 1117
Qy 200 ValTrpSerMetGluSerGlyGlnHisMet-----Proasp----- 211
Db 1118 ATCTGACAAATCGCATCATGTAACACATCGGTGTAACAACATTCCTCCGACTTTGAA 1177
Qy 212 -----LysAlaLysPheValGlyLeuAla----- 220
Db 1178 TTGCTCAGCAGCAAACTCAAAACAGCGGAGCTGATGCTCAACCAACATCACCATTACCA 1237
Qy 221 -----ArgValAlaAlaProGlyAlaIleIleIleValIleThrTrpCysHisArgAsp 238
Db 1238 GACAAACCGCCCGCCAGCGAGGTGATTTAT-----GATCGCTAC 1279
Qy 239 LeuGlyProasp----- 242
Db 1280 ATTTTCCCGACGCTGAACTCACTGCTGCGACCCGTATCAAGCATGCAAGCAAC 1339
Qy 243 -----GlnGlnSerLeuHisPro 248
Db 1340 CGTTTGAAGTCTGCACGACGAAACCTCCGCTTGATTACCAACGACCCCTGACGCG 1399
Qy 249 TrpGlnGlnAspLeuLeuLysLysLleCysAspAlaTyrTyrLeuProAla----- 265
Db 1400 TGGTGGAAACCTCAAGAAATTTGGAGAGACAGTTGAACCTGCGCGTGAACCCACT 1459
Qy 266 -----TrpCysSerThrSerAspTyrVal 273
Db 1460 GCACGACTTGTGGCTGACATGCGAGGTTCGAATGGGATTTGCCACACATTCCTC 1519
Qy 274 LysLeuLeuGlnSerLeuSerLeuGlnAspLleLysSerGluAspTrpSerArgPheVal 293
Db 1520 CAGCTGCACCAAGTACTGCTGTGAA-----CTCGATGACAGGGAAGTCGCGGAGAA 1573
Qy 294 AlaPro-----PheTrpProAlaValIleArg--SerAlaIleThrTrpLysGlyLeu 311
Db 1574 GTTCCTGAAGAAATGTTGTCGACTATCTAAAGAAACAAATGTTCTTTTAAAG----- 1625
Qy 311 rSerLeuLeuSerSerGlyGlnLysThrIleLysGlyAlaLeu 325
Db 1626 -----GTTCCCTAGGGGCGAGAGTTAATTGTAGTAACTC 1661

RESULT 14
US-09-738-626-636
: Sequence 636, Application US/09738626
: Publication No. US20020197605A1
: GENERAL INFORMATION:
: APPLICANT: NAKAGAMA, SATOSHI
: APPLICANT: MIZOGUCHI, HIROSHI
: APPLICANT: ANDO, SEIKO
: APPLICANT: HAYASHI, MIKIRO
: APPLICANT: OCHIAI, KEIKO
: APPLICANT: YOKOI, HARUHIKO
: APPLICANT: TATEISHI, NAOKO
: APPLICANT: SENOH, AKIHITO
: APPLICANT: IKEDA, MASATO
: APPLICANT: OZAKI, AKIO
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-125
: CURRENT APPLICATION NUMBER: US/09/738,626
: PRIOR FILING DATE: 2000-12-18
: PRIOR APPLICATION NUMBER: JP 99/377484
: PRIOR FILING DATE: 1999-12-16
: PRIOR APPLICATION NUMBER: JP 00/159162
: PRIOR FILING DATE: 2000-04-07
: PRIOR APPLICATION NUMBER: JP 00/280988

```


PRIOR FILING DATE: 2000-08-03
 NUMBER OF SEQ ID NOS: 7059
 SOFTWARE: PatentIn ver. 3.0
 SEQ ID NO 636
 LENGTH: 1353
 TYPE: DNA
 ORGANISM: Corynebacterium glutamicum
 US-09-738-626-636

Alignment Scores:
 Pred. No.: 1353
 Score: 139.00
 Percent Similarity: 37.10%
 Best Local Similarity: 22.26%
 Query Match: 7.60%
 Gaps: 12

US-09-857-613a-28 (1-350) x US-09-738-626-636 (1-1353)

53 LeuGlucInLysPro-----LysLysAspAsp 61
 427 ATGGACACACCCCTGATGCGCTAAAGCAGTCAACGGCTAGCATCCAGCAGCTCG 486
 62 LysLysLysLeuGlnLysGlyLleAlaGluPheTyrAspGluSerSergLysLeuTrrpLcu 81
 487 AAATCCCGCAGACAAAGCCATTAGTACACATACGACGCGGCAATGATTTCTACTCC 546
 82 AsnLettRpgLysAspHisMetHis-----GlyPheTyrAspSerAspSerThrVal 99
 547 CTGTTTATGATGATTCATGACCTACCTCGCGCTATTATCCAAAGCCAGACATCA--- 603
 100 SerLysSerAspHisArgLalaAlaGluLysLeuGlnLysLeuArgPheAla 119
 604 AATTTCGAGAGAACCCCAAGAAACAAATACCGCTCATCTTTGAAAACCTGCGTGC--- 660
 120 SerValSerGluGlnArgSerLysTrrpLysSerLysLeuAlaGlyLysGlyLle 139
 661 -----AAGAAGCGGATCGCTCTCTAGACGCGGATCGGTTGG 699
 140 GlyGlySerSerArgTyrLeuAlaLysLysPheGlyValThrSerValGlyLleThrLeu 159
 700 GGAGGCGATGTCGCTC-----GCCGCCAAACACGCGTGTGAAGCCGAGGATTCACGCTG 756
 160 SerProValAlaGlnArgLalaAlaAlaSerAlaLeuAlaAlaGlnGlyLeuAlaAspLys 179
 757 TCTGAACAGCAATGATGATGCGGTCAAGCAGATCAAAACCCCAAGTTTGAAGACCTC 816
 180 ValSerPheGlnValAlaAlaSerAlaLeuGlnInProPheSerAspGlyGlnPheAspLeu 199
 817 GCGGAATTCGCTTCATGATGATTCACCGCATGTTCCA-----GAACATGATTCGATGCG 870
 200 ValTrrpSerMetGluSerGlyGlnHisMet-----ProAspLysAlaLys 214
 871 AACTCAGCAATGCGCATCTTGAACACATCGGTGACACACTATCCGACTACTTTGA 930
 215 PheValGlyLysLeuAlaArgValAlaAlaProGlyAlaLleLleLleValTrrp 234
 931 TTGCTCAGCAGCAAACTCAAAACA-----GCGGAGCTGATGCTC 969
 235 CysHisArgAspLeuGlyProAspGluGlnSerLeuHisProTrrpGluAlaSerLeu 254
 970 AACCAACAGCATCACTCCAGCAACCGCCGCCAC-----GCA 1011
 255 LysLysLleCysAspAlaTyrTyrLeuPro-----AlaTrrpCysSerThrSerAspTyr 272
 1012 GGTGCAATTTATTTGATCGCTACATTTTCCGACAGCTGAACCTCACTGCTGACACCTG 288
 273 ValLysLeuGlnSerLeuSerLeuGlnAspLysSerGluAsp----- 288
 1072 ATCAAGCAATGACAGACAGCGTTTCGAAAGTGTGACAGAAACCTCGCTTGTAT 1131
 289 -----TrrpSerArgPheValAlaProPheTrrpProAlaValLle 301

Db 1132 TACCAAGCAGCCTGTCAGCGCTGTCGAAAACTCAAGAAATTTGGAGAGACAGTT 1191
 QY 302 ArgSerAla 304
 Db 1192 GAACATCGCC 1200

RESULT 15
 US-09-294-093B-5022
 Sequence 5022, Application US/09294093B
 Patent No. US20010051335A1
 GENERAL INFORMATION:
 APPLICANT: Lalqudi, Raghunath, V.
 APPLICANT: Ito, Laura, Y.
 APPLICANT: Sherman, Bradley, K.
 TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
 FILE REFERENCE: PL-0009 US
 CURRENT APPLICATION NUMBER: US/09/294, 093B
 CURRENT FILING DATE: 1999-04-16
 PRIOR APPLICATION NUMBER: 60/082,567
 PRIOR FILING DATE: April 21, 1998
 NUMBER OF SEQ ID NOS: 6207
 SOFTWARE: PERL Program
 SEQ ID NO 5022
 LENGTH: 290
 TYPE: DNA
 ORGANISM: Zea mays
 FEATURE:
 NAME/KEY: misc-feature
 OTHER INFORMATION: Incyte ID No. US20010051335A1 700355594H1
 US-09-294-093B-5022

Alignment Scores:
 Pred. No.: 290
 Score: 128.00
 Percent Similarity: 55.42%
 Best Local Similarity: 33.73%
 Query Match: 10
 Gaps: 1

US-09-857-613a-28 (1-350) x US-09-294-093B-5022 (1-290)

QY 130 LysSerLysLeuAlaSerValGlyCysGlyLleGlyLysSerArgTyrLeuAlaLysLys 149
 20 CAGAAAGTACTGATGATGCGATGCGAATTCGAGCGGCGGACTTTTACATGCTCAAAAC 79
 150 PheGlyAlaThrSerValGlyLleThrLeuSerProValGlnAlaGlnArgAlaAsnAla 169
 80 TATGATGTCATGTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 133
 170 LeuAlaAlaAlaGlnGlyLeuAlaAspLysValSerPheGlnValAlaSerAlaLeuGln 189
 134 ATGGAACGTGCGCATTCGACGCAAGTCTCTGTTGAATTCGAATTCGATTCCTGATTCAC 193
 190 GlnProPheSerAspGlyGlnPheAspLeuValTrrpSerMetGluSerGlyGlnHisMet 209
 194 AAGGATTACCAAGAAATATCTTTTGAACATCATCATACGCGGTGACACACATCCTTCACATA 253
 210 ProAspLys 212
 254 CAGCAAG 262

Search completed: March 19, 2003, 09:06:49
 Job time : 100 secs.

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Run on: March 19, 2003, 06:53:46 ; Search time 1289 Seconds
(without alignments)

4397.533 Million cell updates/sec

Title: US-09-857-613A-28
1930

Sequence: 1 MATVVRIPITSCIHIFRS.....IEGYKDLIKFAITCRKPE 350

Scoring table:

BLOSUM62		
Xgapop	10.0	Xgapext 0.5
Ygapop	10.0	Ygapext 0.5
Fgapop	6.0	Fgapext 7.0
Delop	6.0	Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match 0%
	Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```
-MODE=iframexp2n;model DEV=fast
-cg/cg2.1/JSPTO_spool/US098576113/runat.12032003.100036.6027/app.query.fasta.1.51
-DB=ESTR-fastap SUFIX=rs1-MINMATCH=0.1 -LOCP=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRY-blosum62 -TRANS=humana40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=1 -MAXLEN=200000000
-USER=US098576113 @cgn.1.1.1456 @runat.12032003.100036.6027 -NCPU=6 -ICPU=3
-NO_XLIFY -NO_MMAP -LARGESEUER -NEG_SCORE=0 -WAIT -LONLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPEXT=0.5 -FGAPEXT=6 -FGAPEXT=7
-YGAPEXT=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database ::

```

E87: *
1:  em_estba: *
2:  em_estbhum: *
3:  em_estlin: *
4:  em_estmu: *
5:  em_estrov: *
6:  em_estropl: *
7:  em_estro: *
8:  em_hlc: *
9:  qb_est1: *
10: qb_est2: *
11: qb_hlc: *
12: qb_est3: *
13: qb_est4: *
14: qb_est5: *
15: em_estfun: *
16: em_estcom: *
17:  gd_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_man: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27:  em_gss_rod: *

```

Result	Score	Query	Match	length	DB	ID	Description
No.							
1	1136.5	62.1	1304	11	AY109799	AY109799 Zea mays	
2	1029	56.2	599	14	BO299297	BO299297 sa045f11..	
3	1024	56.0	669	10	BE311080	BE311080 so05a05.Y	
4	980	53.6	567	14	BO273543	BO273543 sa02a4e1..	
5	975	53.3	570	14	BO786302	BO786302 sa066c02.	
6	971.5	53.1	778	12	BM678630	BM678630 ESTf473i68	
7	964	52.7	12	BS127606	BS127606 sa166b10..		
8	957	52.3	562	13	BM527813	BM527813 sa110d11..	
9	946	51.7	556	13	BI469324	BI469324 sa109h01..	
10	931	50.9	704	13	BI469281	BI469281 sa09h01..	
11	922	50.4	831	12	Bj288530	Bj288530 Zm03_01d1	
12	899	49.1	664	13	Bj463750	Bj463750 Bj463750	
13	884	48.3	664	14	Bj233302	Bj233302 sa033e12..	
14	878	48.0	515	14	BO213302	BO213302 sa033e12..	
15	869	47.5	628	10	AM566072	AM566072 s197e09..Y	
16	855	46.7	581	9	AI468374	AI468374 EST245696	
17	844	46.1	509	12	BI321569	BI321569 sa012h01..	
18	840	45.9	484	13	BI311583	BI311583 sa011s05..	
19	836	45.7	708	14	BO801716	BO801716 WHE2817-G	
20	824	45.0	503	12	BG790250	BG790250 sa067a12..	
21	810	44.3	649	13	BI953869	BI953869 HVSMEK001	
22	807	44.1	707	12	BG416854	BG416854 HVSMEK001	
23	794	43.4	691	14	BQ116842	BQ116842 EST602418	
24	788	43.1	540	10	BE331113	BE331113 so09306..Y	
25	777	42.5	517	12	BG046436	BG046436 sa054a02..	
26	775	42.3	587	10	BE340311	BE340311 EST344382	
27	769	41.7	627	12	BG522228	BG522228 19-89 Ste	
28	763	41.0	708	12	BG416864	BG416864 HVSMEK001	
29	759	41.5	439	12	BE804492	BE804492 sr080d07.Y	
30	754	41.2	444	13	BM094790	BM094790 saJ21d02..	
31	734.5	40.1	553	10	AM774108	AM774108 EST333338	
32	728	39.8	553	12	BF631792	BF631792 NF008D03D	
33	720	39.3	587	10	BE591261	BE591261 WHE1655-1	
34	710	38.8	661	13	Bj447900	Bj447900 BJ447900	
35	707	38.6	733	12	BF051091	BF051091 EST3436266	
36	685.5	37.5	533	9	AI1489112	AI1489112 EST247451	
37	670	36.6	517	10	BE607043	BE607043 WHE0915-H	
38	665	36.3	484	10	AM611931	AM611931 EST314005	
39	662.5	36.2	730	12	BG523936	BG523936 37-16 Ste	
40	660	36.1	639	10	AY945401	AY945401 AY945401	
41	658	36.0	491	9	AI1489517	AI1489517 EST244238	
42	657	35.9	531	10	BE426891	BE426891 WHE0333-C	
43	656	35.8	452	10	AM099184	AM099184 sd35h03.Y	
44	650	35.5	623	12	BG447567	BG447567 NF004H015	
45	642	35.1	449	12	BG725508	BG725508 sae38c12..	

ALIGNMENTS

RESULT 1				
AY109799				
LOCUS	AY109799	1304 bp	mRNA	linear
DEFINITION	Zea mays CL364_1 mRNA sequence.			
ACCESSION	AY109799			
VERSION	AY109799.1	GI:21213657		
KEYWORDS	HTC.			
SOURCE	Zea mays.			
ORGANISM	Zea mays.			
REFERENCE	Enkaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.			
AUTHORS	1 (bases 1 to 1304) Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.			


```

/organism="Glycine max"
/db.xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl054-7197"
/clone_lib="Gm-cl054"
/tissue_type="Leaf, 3 week old, greenhouse grown"
/lab_host="DH10B"
/notes="Vector: pBluescript II SK-, Site.1: EcoRI; Site.2:
XhoI; The Harosoy NIL was constructed and seed was
provided by Dr. J. Specht, University of Nebraska
(Shoemaker and Specht, 1995). The cDNA library was
constructed from mRNA isolated from leaf tissue at various
developmental stages of 3 week old greenhouse grown
plants. Complementary DNA was synthesized from mRNA using
a primer consisting of a poly(dT) sequence with a XhoI
restriction site and a 3' anchor. EcoRI adapters were
ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into DH10B host cells (GibcoBRL). This
library was constructed in cooperation with Dr. Paul
Keim's laboratory at Northern Arizona University."
BASE COUNT      139 a      144 c      165 g      151 t
ORIGIN

```

```

Alignment Scores:
Pred. No.:      6,27e-96      Length:      599
Score:          1029.00      Matches:      199
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     56,23%      Indels:      0
DB:              14          Gaps:      0

```

US-09-857-613a-28 (1-350) x BQ299297 (1-599)

```

QY 6 ArgIleProThrIleSerCysIleHisIleHisThrPheArgSerGlnSerProArgThr 25
    |||||||
DB 3 AGATATCCCAACATTCATGATCCATCCATCCACATCCGCTCCGATCCCAATCCCTCCACAT 62
QY 26 PheAlaArgIleArgValAlaGlyProArgSerTrpAlaProIleArgAlaSerAlaIleSer 45
    |||||||
DB 63 TTCGCCAGATCCGGCTCGGAGACCAGCTGGTGGCTTCCTATTCGGGATCCGACGAGAC 122
QY 46 SerGluArgGlyGluIleValIleGluGlnIleProLysProLysLysAspLysLysLeu 65
    |||||||
DB 123 TCGGAGAGAGGGAGATGATATTGGAGCAGAACCCGAAAGAGATGACAAGAAAGAGCTG 182
QY 66 GlnLysGlyIleAlaGluPheTyrAspGluSerSerGlyLeuTrpGlnIleTrpGly 85
    |||||||
DB 183 CAGAAAGGAAATCGACAGATTTCACGACGAGCTCTTCGCTTATGGCAACAATTCGGGC 242
QY 86 AspHisMetHisIleGlyPheTyrAspSerAspSerThrValSerLeuSerAspHisArg 105
    |||||||
DB 243 GACCAATCAGACATCGGCTTTATGATCGATTCACATGCTTCCTTCGGATCATCTCT 302
QY 106 AlaAlaGlnIleArgMetIleGlnIleSerLeuArgPheAlaSerValSerGluArg 125
    |||||||
DB 303 GCTGCTCAGATCCGATGATCCAAAGATCTCTTCGCTTCCTGCTTCGAGAGCGCT 362
QY 126 SerLysTrpProLysSerIleValAspValAlaGlyCysGlyIleGlyIleSerSerArgTyr 145
    |||||||
DB 363 AGTAAATGCCCCAAGAGATGATGATGATGGTGGCATGGGCGAGCTCTAGATAC 422
QY 146 LeuAlaLysLysPheGlyValAlaThrSerValGlyIleThrLeuSerProValGlnIleArg 165
    |||||||
DB 423 CTGGCCAGAAATTTGGAGCAACAGCTAGCATCTCTAGTCCCTGTTCAAGCTCAA 482
QY 166 ArgAlaAsnAlaLeuAlaAlaAlaGlnIleGluAlaAspLysValSerPheGlnValAla 185
    |||||||
DB 483 AGAGCAAAATGCTCTGCTGCTCAAGATGGCTGATAGGTTCTCTTCAGGTTGCT 542
QY 186 AspAlaLeuGlnGlnProPheSerAspGlyGlnPheAspLeuValTrpSerMetGlu 204
    |||||||

```

DB 543 GACGCTACAGCAACCATTCCTGAGCGCCAGTTGATCTGTGTGTCATGAG 599

```

RESULT 3
LOCUS BE331080
DEFINITION BE331080 669 bp mRNA linear EST 04-DEC-2001
          sc95a05.y1 Gm-cl041 glycine max cDNA clone GENOME SYSTEMS CLONE ID:
          Gm-cl041-1065 5' similar to TR:09ZSK1 09ZSK1 GAMMA-T0COPHEROL
          METHYLTRANSFERASE. ; mRNA sequence.
ACCESSION BE331080
VERSION BE331080.1 GI:9204856
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eustoids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
          Glycine.
REFERENCE 1 (bases 1 to 669)
          Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Corvett,V., Rhanna
          ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
          Wylie,T., Underwood,K., Steptoe,M., Rheising,B., Allen,M., Bowers
          ,Y., Peterson,B., Swaller,T., Giddons,M., Pape,D., Harvey,N., Schurk
          ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
          ,R., Waterston,R. and Wilson,R.
          Public Soybean EST Project
          Unpublished (1999)
          Contact: Shoemaker R/Public Soybean EST Project
          Public Soybean EST Project
          Washington University School of Medicine
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: est@watson.wustl.edu
          This clone is available through: Resgen, Invitrogen Corp. 2130
          South Memorial Parkway Huntsville, AL 35801 For further information
          call: (800)-533-4363 or contact via email: c@resgen.com
          High quality sequence stop: 461.
          Location/Qualifiers
            1..669
              /organism="Glycine max"
              /db.xref="taxon:3847"
              /clone="GENOME SYSTEMS CLONE ID: Gm-cl041-1065"
              /clone_lib="Gm-cl041"
              /tissue_type="Senescing leaves, mature plant, greenhouse
              grown"
              /lab_host="DH10B"
              /note="Vector: pTrpT3pac (Pharmacia); Site.1: EcoRI;
              Site.2: HindIII. This library was constructed from mRNA
              isolated from senescing leaf tissue of mature greenhouse
              grown plants of the cultivar Williams. Complementary DNA
              was synthesized from mRNA using a 3' anchored poly(dT)
              primer. EcoRI adapters were ligated to the blunt-ended
              cDNA fragments followed by digestion with EcoRI and
              HindIII. The cDNA fragments were directionally cloned
              into the EcoRI-HindIII restriction site of the pTrpT3-pac
              vector. The ligated cDNA fragments were transformed into
              DH10B host cells. This library was constructed by Dr.
              Randy Shoemaker."

```

BASE COUNT 154 a 156 c 184 g 174 t 1 others

```

ORIGIN
Alignment Scores:
Pred. No.:      2.37e-95      Length:      669
Score:          1024.00      Matches:      200
Percent Similarity: 97.10%      Conservative: 1
Best Local Similarity: 96.62%      Mismatches: 6
Query Match:     55.96%      Indels:      0
DB:              10          Gaps:      0

```

US-09-857-613a-28 (1-350) x BE331080 (1-669)

```

QY 1 MetAlaThrValAlaArgIleProThrIleSerCysIleHisIleHisThrPheArgSer 20
    |||||||

```

This clone is available through: Resgen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: 1-800-533-4363 or contact: cu@resgen.com web site: www.resgen.com

Seq primer: -40RP from Gibco
High quality sequence stop: 420.

FEATURES

source

```

source
1. 20/
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-c1054-5206"
/clone.lib="Gm-c1054"
/tissue_type="leaf, 3 week old, greenhouse grown"
/lab_host="DH10B"
/ab_host="DH10B"
/notes="Vector: pBluescript II SK+; Site.1: EcoRI; Site.2:
XhoI; The Harosoy NIL was constructed and seed was
provided by Dr. J. Specht, University of Nebraska
(Shoemaker and Specht, 1995). The cDNA library was
constructed from mRNA isolated from leaf tissue at various
developmental stages of 3 week old greenhouse grown
plants. Complementary DNA was synthesized from mRNA using
a primer consisting of a poly(dT) sequence with a XhoI
restriction site and a 3' anchor. EcoRI adapters were
ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into DH10B host cells (GibcoBRL). This
library was constructed in cooperation with Dr. Paul
Winters, Laboratory at Northern Arizona University."

```

BASE COUNT ORIGIN	133 a	140 c	144 s
1	100	100	100
2	100	100	100
3	100	100	100
4	100	100	100
5	100	100	100
6	100	100	100
7	100	100	100
8	100	100	100
9	100	100	100
10	100	100	100
11	100	100	100
12	100	100	100
13	100	100	100
14	100	100	100
15	100	100	100
16	100	100	100
17	100	100	100
18	100	100	100
19	100	100	100
20	100	100	100
21	100	100	100
22	100	100	100
23	100	100	100
24	100	100	100
25	100	100	100
26	100	100	100
27	100	100	100
28	100	100	100
29	100	100	100
30	100	100	100
31	100	100	100
32	100	100	100
33	100	100	100
34	100	100	100
35	100	100	100
36	100	100	100
37	100	100	100
38	100	100	100
39	100	100	100
40	100	100	100
41	100	100	100
42	100	100	100
43	100	100	100
44	100	100	100
45	100	100	100
46	100	100	100
47	100	100	100
48	100	100	100
49	100	100	100
50	100	100	100
51	100	100	100
52	100	100	100
53	100	100	100
54	100	100	100
55	100	100	100
56	100	100	100
57	100	100	100
58	100	100	100
59	100	100	100
60	100	100	100
61	100	100	100
62	100	100	100
63	100	100	100
64	100	100	100
65	100	100	100
66	100	100	100
67	100	100	100
68	100	100	100
69	100	100	100
70	100	100	100
71	100	100	100
72	100	100	100
73	100	100	100
74	100	100	100
75	100	100	100
76	100	100	100
77	100	100	100
78	100	100	100
79	100	100	100
80	100	100	100
81	100	100	100
82	100	100	100
83	100	100	100
84	100	100	100
85	100	100	100
86	100	100	100
87	100	100	100
88	100	100	100
89	100	100	100
90	100	100	100
91	100	100	100
92	100	100	100
93	100	100	100
94	100	100	100
95	100	100	100
96	100	100	100
97</			

Alignment Scores:

Alignment Scores:	
pred. No.:	6.56e-91
Score:	980.00
Percent Similarity:	99.47%
Best Local Similarity:	99.47%
Query Match:	53.55%
na:	14
length:	567
Matches:	187
Conservative:	0
Matches:	1
Indels:	0
Gaps:	0

US-09-857-613A-28 (1-350) X BQ273543 (1-567)

QY	89	IIHSHLSGLYPHEtyrAspSerAspSerthrValSerLeuSerAspHisArgAlaAlaGln	108
Db	3	CACCAATGGCTTTATGACTGGATGCCACTGTCTTCGCTTGCATCATCGTCTGCTCAG	62
QY	109	IIeaRMeLIleGlnGluSerLeuArgPheAlaSerValSerGluGlnuArgSerLysrTP	128
Db	63	ATCCGAATATCCAGACAGCTCTTCGCTTCCCTCTGTTCTGAGAGAGCCTACTGAATG	122
QY	129	ProLysSerIleValAspValGlyGlySerGlyIleGlyGlySerSerArgTyrLeuAlaLys	148
Db	123	CCCAAGAGATATATGATATTTGGGTGGGTGCATAGCTGGCAGCCTCAGATACCTGGCCAG	182
QY	149	LysPheGlyValathrSerValGlyIlethrLeuSerProValGlnAlaGlnArgAlaAsn	168
Db	183	AAATTTGAGACACCAAGTGAAGCATCACTGATCCTCGTTCCAGCTCAAGCTCAAGAGCAAT	244
QY	169	AlaLeuAlaAlaAlaGlnGlyLeuAlaAspLysValSerPheGlnValAlaAspAlaLeu	186
Db	243	GCTCTGCTGCTGTCMAAGATTTGCGTGTAAAGTTTCTTTCAAGTTCTCAAGCTCTCA	304
QY	189	GlnGlnProPheSerAspGlyGlnPheAspLeuValTyrSerMetGluSerGlyGluHis	204
Db	303	CAGCAACACTTCTTGACGGCGCAAGTTTGAATCTGTGTGTCTCATGAGAGCTGGAAGCAT	364
QY	209	MetProAspLysAlaLysPheValGlyGlnLeuAlaArgValAlaAlaProGlyIleIle	224
Db	363	ATGCCGTGCAAAAGCTAACTTTGTTGGAGAGTTAGCTGCGGTACACACCCAGCTGCCACT	424
QY	229	IleIleIleValAlaThrTyrCysHisArgAspLeuGlyProAspGluGlnSerLeuHisPro	244

Db 423 ATATATAGTACATGGTGGACAGGAGTCTGGCCCTGACGACAAATCTTACATCA 482

QY 249 TTPGLUGLNASPLEULELysLYsILECysAspAlaTYrTYrLeuProAlaTrpSer 268
|||||

Db 483 TGGAGCAAGATCTCTTAACAAAGATTGGCATGATTAATCACTCCCTGGTGTCA 542
|||||

QY 269 ThisSerAspTYrValLYsLeuLeu 276
|||||

Db 543 ACTCTGATTATGTTAGTGTCTC 566
|||||

RESULT 5
B0786302

LOCUS DEFINITION B0786302 570 bp mRNA linear EST 26-JUL-2002
sag66c02.y1 Gm-cl076 Glycine max cDNA clone SOYBEAN CLONE ID:
Gm-cl076-4708 5' similar to TR:09XHP9 09XHP9 GAMA-TCCOPHEROL
METHYLTRANSFERASE. [1] mRNA sequence.

ACCESSION B0786302
VERSION B0786302.1 GI:21994774

KEYWORDS
SOURCE soybean.
ORGANISM Glycine max
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots;
Rosidae: eustosids I: Fabales; Fabaceae: Papilionoideae: Phaseoleae;
Glycine.
1 (bases 1 to 570)
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna
, A., Bolla, B., Merritt, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
, T., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuck
, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., Mccann
, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Resgen, Invitrogen Corp. 2130
South Memorial Parkway Huntville, AL 35801 For further information
call: (800)-533-4363 or contact: cவுeresgen.com web site:
www.resgen.com
Seq primer: -40RP from Gibco
High quality sequence stop: 422.
Location/Qualifiers
1. 570
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl076-4708"
/clone_id="Gm-cl076"
/tissue_type="wounded cotyledons"
/dev_stage="11 day old seedlings"
/lab_host="DH10B"
/note="Vector: pBluescript II SK+, Site_1: EcoRI, Site_2:
XhoI. The cDNA library was constructed from mRNA isolated
from 11 day old seedlings treated with that were treated
with 2 ugs/ml of a crude glucan elicitor preparation
isolated from the mycelial walls of Phytophthora sojae.
The library was prepared using the Stratagene pBluescript
II SK(+) library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with a blunt restriction site. EcoRI
adaptors were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA fragments were
directionally cloned into the EcoRI-XhoI restriction site
of the pBluescript vector. The ligated cDNA fragments were
transformed into E.coli ElectroMax DH10B host cells. Plant
material was provided by Michael G. Hahn (Complex
Carbohydrate Research Center, University of Georgia) and
the library was constructed by Anu Khanna (Lila Vodkin Lab

BASE COUNT 156 a 123 c 136 g 155 t
ORIGIN
Alignment Scores:
Pred. No.: 2,16e-90 Length: 570
Score: 975.00 Matches: 185
Percent Similarity: 99.47% Conservative: 1
Best Local Similarity: 98.93% Mismatches: 1
Query Match: 53.28% Indels: 0
DB: 14 Gaps: 0

US-09-857-613a-28 (1-350) x B0786302 (1-570)

QY 164 ALaGlnArGAlaAsnAlaLeuAlaAlaGlnGlyLeuAlaAspLYsAlaSerPheGln 183
|||||

Db 1 GCTTAAGAGCAAAATGCTCTGCTGCTGCTCAAGAGATGGCTGATTAAGCTTCCTTCAG 60
|||||

QY 184 ValAlaAspAlaLeuGlnGlnInProPheSerAspGlyInPheAspLeuValTrpSerMet 203
|||||

Db 61 GTTGCTAGCGCTCTACAGCAACCATTTCTTGACGGCCAGTTTATCTGCTGCTCCATG 120
|||||

QY 204 GlnSerGlyGlnHisMetProAspLYsAlaLYsPheValGlyLeuLeuAlaArgValAla 223
|||||

Db 121 GAGAGTGGAGAGCATATGCTGCACAAAGCTAATTGTTGGACAGTACCTCGGCTAGCA 180
|||||

QY 224 AlaProGlyAlaLeuLeuLeuLeuLeuValThrTrpCysHisArgAspLeuGlyProAspGlu 243
|||||

Db 181 GCACCAAGGTCCACTATTAATTAATGATGATGCTGCCACAGGATCTGGCCCTGACGAA 240
|||||

QY 244 GlnSerLeuHisProTrpGlnAspLeuLeuLYsILECysAspAlaTYrTYrLeu 263
|||||

Db 241 CAATCCTTACATCCATGAGGAGCAAGATCTCTTAAGAAAGATTTGCGATGATTAACCTC 300
|||||

QY 264 ProAlaTrpCysSerThrSerAspTYrValLYsLeuLeuGlnSerLeuSerLeuGlnAsp 283
|||||

Db 301 CCGCGCTGGTGCACCACTTCTGATTAATGTTAAGTCTCCATCCCTGCACCTCAGGC 360
|||||

QY 284 ILeLYsSerGlnAspTrpSerArgPheValAlaTrpPheTrpProAlaValIleArgSer 303
|||||

Db 361 ATCAAGTCAAGAAAGATTGGCTGCTGCTTGTGCTCCATTTTGGCCAGCAGATGATACGCTCA 420
|||||

QY 304 AlaPheThrTrpLYsGlyLeuSerSerLeuSerSerGlyGlnLYsThrIleLYsGly 323
|||||

Db 421 GCCTTCACATGGAAGGCTGTACTTCACTTCACTGACAGCAGTGCACAAAGACGATTAAGAGA 480
|||||

QY 324 AlaLeuAlaMetProLeuMetIleGlnGlyTYrLYsAspLeuIleLYsPheAlaIle 343
|||||

Db 481 GCTTGGCTATGCCATTGATGATACAGGAGATACAAAGAAAGATTAATTAAGTTGCCATTC 540
|||||

QY 344 ILeThrCysArgLYsProGlu 350
|||||

Db 541 ATTACATGTCGAAACCTGAA 561
|||||

RESULT 6
B6127606

LOCUS DEFINITION B6127606 778 bp mRNA linear EST 31-JAN-2001
EST473168 tomato shoot/meristem Lycopersicon esculentum cDNA clone
CTOF17C24 5' sequence, mRNA sequence.

ACCESSION B6127606
VERSION B6127606.1 GI:12627710

KEYWORDS
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots;
Asterales: euasterids I: Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 778)
van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Utterback, T.,
Hansen, C., Romling, C. and Tanksley, S.
Generation of ESTs from tomato shoot/meristem tissue
unpublished (2001)

ORIGIN

Alignment Scores: 2.84e-89 Length: 554
 Pred. No.: 964.00 Matches: 183
 Score: 99.468
 Percent Similarity: 99.468
 Best Local Similarity: 99.468
 Query Match: 52.688
 DB: 13 Gaps: 0

US-09-857-613a-28 (1-350) x BM527813 (1-554)

OY 68 GYIIEAIGLuphEYrAspGluSerSerGlyLeuTTPGluAsnIleTTPGlyAspHis 87
 DB 2 GGAATCGAGAGTTTACAGACAGCTTCTGCTTATGAGAACATTTGGGCGACAC 61
 OY 88 MethIsHsIGlypHeTyrAspSerSerThValSerLeuSerAspHisrGluAla 107
 DB 62 ATGACCATGGCTTTATGACTCGGATTCACCTGTTTGGCTTCGATCCTGCTCT 121
 OY 108 GlnIleArgMetIleGlnGluSerLeuArgPheAlaSerValSerGluGluArgSerLys 127
 DB 122 CAGATCCGAATGATCCAGAGCTCTTCTGCTTCTGCTTCTGAGAGCTAGTAA 181
 OY 128 TTPProLysSerIleValAspValGlyGlySerGlyIleGlySerSerArgTyrLeuAla 147
 DB 182 TGGCCCAAGATTAATGATGATGTTGGTGGCATAGCTGCACCTCTAGATACCTGGCC 241
 OY 148 LysLysPheGlyAlaThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAla 167
 DB 242 AAGAAATTTGGAGCAGACAGCTGAGCATCTGATCTCTGTTCAAGCTCAAGACCA 301
 OY 168 AsnAlaLeuAlaAlaIleGlnGlyLeuAlaAspLysValSerPheGlnValAlaAspAla 187
 DB 302 AATGCTCTTGGCTGCTCAAGAGATGCTGATAGAGTTTCTTCAAGCTGCTGAGCCT 361
 OY 188 LeuGlnGlnProPheSerAspGlyGlnPheAspLeuValTTPSerMetGluSerGlyGlu 207
 DB 362 CTACAGACACATTTCTGAGCGCCAGCTTGTATGCTGCTGCTCAGATGAGAGAG 421
 OY 208 HisMetProAspLysAlaLysPheValGlyGluLeuAlaArgValAlaAlaProGlyAla 227
 DB 422 CATATGCTGACCAAGCTAAGTTTGTGGAGAGTTTACCTCGGATGACACCGAGTGC 481
 OY 228 IleIleIleIleValIleTTPCysHisArgAspLeuGlyProAspGluGlnSerLeuHis 247
 DB 482 ACTATATTAATAGTACATGATGTCACAGAGATCTTGCCCTGACGAAACATCTTACAT 541
 OY 248 ProTTPGluGln 251
 DB 542 CCATGGAGCA 553

RESULT 8

BI469324

LOCUS

DEFINITION

ACCESION

VERSION

KEYWORDS

SOURCE

ORGANISM

soybean.

glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

glycine.

1 (bases 1 to 562)

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna,

A., Bolla, B., Marita, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,

Wille, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers,

Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schuck,

R., Rlter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann

TITLE
 JOURNAL
 COMMENT
 'R., Waterston, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu

FEATURES
 Source
 1..562
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl053-2854"
 /clone_lib="Gm-cl053"
 /tissue_type="Whole seedling, 3 week old, greenhouse grown"
 /lab_host="DH10B"
 /note="Vector: pBluescript II SK+; Site.1: EcoRI; Site.2: XhoI; The Harosoy NIL was constructed and seed was provided by Dr. J. Specht, University of Nebraska (Shoemaker and Specht, 1995). The cDNA library was constructed from mRNA isolated from whole seedlings of 3 week old greenhouse grown plants. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site and a 3' anchor. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in cooperation with Dr. Paul Keim's Laboratory at Northern Arizona University."

BASE COUNT 130 a 136 c 153 g 143 t

ORIGIN

Alignment Scores: 1.52e-88 Length: 562
 Pred. No.: 957.00 Matches: 186
 Score: 99.478
 Percent Similarity: 99.478
 Best Local Similarity: 99.478
 Query Match: 52.308
 DB: 13 Gaps: 0

US-09-857-613a-28 (1-350) x BI469324 (1-562)

OY 11 SerCysIleHisIleHisThrPheArgSerGlnSerProArgThrPheAlaArgIleArg 30
 DB 2 TCATGACATCCACATCCACAGCTTCGTTCCCAATCCCTCCACTTTCGCCAATCCGG 61
 OY 31 ValGlyProAlaGlySerTTPAlaProIleArgAlaSerAlaIleSerSerGluArgGlyGlu 50
 DB 62 GTCGAGCCACAGCTGCTGCTCTATTCGCGCATCCGACGACGACCTCGAGAGAGGAG 121
 OY 51 IleValLeuGlnGlnLysProLysLysAspAspLysLysLysLysLysLysGlyIleAla 70
 DB 122 ATAGTATTGGAGCAAGAGCCCAAGAGATGACAAAGAGAGCTGCAGAGAGATGCA 181
 OY 71 GluPheTyrAspGluSerSerGlyLeuTTPGluAsnIleTTPGlyAspHisMetHis 90
 DB 182 GAGTTTACGACGAGCTCTTGGCTTATGAGAGAACATTTGGGCGACACATCCACAT 241
 OY 91 GlyPheTyrAspSerSerThrValSerLeuSerSerAspHisArgAlaIleArg 110
 DB 242 GCGTTTATAGCTCGATTCACATGTTTCTTCCGATCATGCTGCTCAGATCCGA 301
 OY 111 MetIleGlnSerLeuArgPheAlaSerValSerGluGluArgSerLysTTPProLys 130


```

|||||
Db 302 ATGATCCAGAGCTCTCCCTTCCCTGCTTCTGAGAGCTAGTAAATGCCCAAG 361
|||
Qy 131 SerileValaspValGlyCysGlyIleGlyGlySerSerArgTyrLeuAlaLysLysPhe 150
|||
Db 362 AGTATAGTATGATGTTGGTGGTGCATAGTGGCAGCTCTGATGATCCGCGCAAGAAATTT 421
|||
Qy 151 GlyValThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaAspAlaLeu 170
|||
Db 422 GGACCAACCGAGTGAAGCATCCTCTGCTTCAAGCTCAAGAGCAAAATGCTCTT 481
|||
Qy 171 AlaAlaIleGlnGlyLeuAlaAspLysValSerPheGlnValAlaAspAlaLeuGlnGln 190
|||
Db 482 GCTGCTGCTCAAGGATGCTGATAGGCTTCTTTCAGTTGCTGACGCTCTACAGCA 541
|||
Qy 191 ProPheSerAspGlyGlnPhe 197
|||
Db 542 CCATTCCTCTGAGCGCCAGTTT 562
|||
RESULT 9 556 bp mRNA linear EST 30-NOV-2001
LOCUS B1469281
DEFINITION B1469281.1 GI:15285390
ID: Gm-cl053-3025 5', similar to TR:Q9XIP9 Q9XIP9 GAMMA-TOCOPHEROL
METHYLTRANSFERASE. [1] ;, mRNA sequence.
ACCESSION B1469281
VERSION B1469281.1 GI:15285390
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 556)
Shoemaker, R., Kelm, P., Vodkin, L., Erpelting, J., Corvett, V., Khanna
, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wylie, T., Underwood, K., Steploe, M., Theising, B., Allen, M., Bowers
, K., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available through: Resgen, Invitrogen Corp. 2130
South Memorial Parkway Hunttsville, AL 35801 For further information
call: (800)-533-4363 or contact via email: cou@resgen.com
High quality sequence stop: 422.
Location/Qualifiers
1..556
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl053-3025"
/clone_lib="Gm-cl053"
/tissue_type="Whole seedling, 3 week old, greenhouse
grown"
/lab_host="DH10B"
/note="vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
XhoI; The Harosoy NIL was constructed and seed was
provided by Dr. J. Specht, University of Nebraska
(Shoemaker and Specht, 1993). The cDNA library was
constructed from mRNA isolated from whole seedlings of 3
week old greenhouse grown plants. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with a XhoI restriction site and a 3'
anchor. EcoRI adapters were ligated to the blunt-ended
cDNA fragments followed by XhoI digestion. The cDNA

```

```

fragments were directionally cloned into the EcoRI-XhoI
restriction site of the pBluescript vector. The ligated
cDNA fragments were transformed into DH10B host cells
(GibcoBRL). This library was constructed in cooperation
with Dr. Paul Keim's laboratory at Northern Arizona
University."
BASE COUNT 129 a 135 c 152 g 140 t
ORIGIN
Alignment Scores:
Pred. No.: 2,04e-87 Length: 556
Score: 946.00 Matches: 184
Percent Similarity: 99.46% Conservative: 0
Best Local Similarity: 99.46% Mismatches: 1
Query Match: 51.69% Indels: 0
Gaps: 0
DB: 13
US-09-857-613a-28 (1-350) x B1469281 (1-556)
Qy 11 SerCysIleHisIleHisThrPheArgSerGlnSerProArgThrPheAlaArgIleArg 30
|||
Db 2 TCATGCATCCACATCCACACGCTTCCTCCCAATCCCTCGCACTTCCGCAAGATCCG 61
|||
Qy 31 ValGlyProArgSerTyrPheAlaProIleArgAlaSerAlaIleSerSerGluArgGlyGlu 50
|||
Db 62 GTCGACCCAGAGTCGTGGGCTCTATTCCGGCATCGCCAGCAGCTCGAGAGAGGGAG 121
|||
Qy 51 IleValLeuGlnGlnLysProLysLysAspAspLysLysLysLeuGlnLysGlyIleAla 70
|||
Db 122 ATGATGATGGAGCAGAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 181
|||
Qy 71 GluPheTyrAspGlnSerSerGlyLeuTyrPheLysAsnIleTyrGlyAspHisMetHisHis 90
|||
Db 182 GAGTTTACGACGAGCTTCTTCGCTTATGAGAGAAATTGGCGGCGACCATGCACCAT 241
|||
Qy 91 GlyPheTyrAspSerAspSerThrValSerLeuSerAspHisArgAlaIleGlnIleArg 110
|||
Db 242 GCGTTTATGATCGATGATTCACATGCTTTCGCTTGGATCATGCGTCTCAGATCCGA 301
|||
Qy 111 MetIleGlnIleSerLeuArgPheAlaSerValSerGluGluArgSerLysTyrProLys 130
|||
Db 302 ATGATCCAGAGATCTCTGCTTCCCTGCTTCTGCTTCTGAGAGAGAGAGAGAGAGAG 361
|||
Qy 131 SerileValaspValGlyCysGlyIleGlyGlySerSerArgTyrLeuAlaLysLysPhe 150
|||
Db 362 ACTATAGTGAATGTTGGTGGTGCATAGTGGCAGCTCTGATGATCCGCGCAAGAAATTT 421
|||
Qy 151 GlyValThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaAspAlaLeu 170
|||
Db 422 GGAGCAACCGAGTGAAGCATCCTCTGCTTCAAGCTCAAGAGCAAAATGCTCTT 481
|||
Qy 171 AlaAlaIleGlnGlyLeuAlaAspLysValSerPheGlnValAlaAspAlaLeuGlnGln 190
|||
Db 482 GCTGCTGCTCAAGGATGCTGATAGGCTTCTTTCAGTTGCTGACGCTCTACAGCA 541
|||
Qy 191 ProPheSerAspGly 195
|||
Db 542 CCATTCCTCTGAGCGC 556
|||
RESULT 10 704 bp mRNA linear EST 09-APR-2002
LOCUS B1288530
DEFINITION B1288530 Y. Ogihara unpublished cDNA library, Wh. SL Triticum
aestivum cDNA clone whs114p22 5', mRNA sequence.
ACCESSION B1288530
VERSION B1288530.1 GI:20106970
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae
; Triticaceae; Triticum.
1 (bases 1 to 704)

```

AUTHORS	Ogihara, Y. and Murai, K.
TITLE	Expressed genes in <i>Triticum aestivum</i>
JOURNAL	Unpublished (2002)
COMMENT	Contact: Tadasu Shin-i

FEATURES	Location/Qualifiers
SOURCE	1. .704

Alignment Scores:	
Pred. No.:	9,74e-86
Score:	931.00
Percent Similarity:	85.46%
Best Local Similarity:	76.21%
Query Match:	50.87%
DB:	13
	0
Length:	704
Matches:	173
Conservative:	21
Mismatches:	33
Indels:	0
Gaps:	0

US-09-857-613A-28 (1-350) x BJ288530 (1-704)

QY	83	ILERPPGLYASPHISMETHSHISGLPHEHYRXPSPERKSPERTHVALSERLEUSER	102
Db	22	ATCGGGGGCGAGCAATGCCACACGCGTTCTTACGACATCCGGCAGCGCGCTTCATATGCC	81
QY	103	ASPHISARGALAALAGINILEARGMETILEGLINLSERLEUARGPHEALASERVALSER	122
Db	82	GACCAACGCGCGCGCGCGAGATCCGATGATCGAGAGAGGCGCTCCGCTTCGGCGCGTCGCC	141
QY	123	GLIULIARGSERLYSTRPPROPYLSERILEVALASPYALJLYCYSGLYILEGLYGLYSER	142
Db	142	GAGCATCCGAGAAACAAACCGAAACCAATGTTGATGTCGGAATGCCGAATCGGTGGAGC	201
QY	143	SERRGTYRILEUALALYSLYSPHEGLIATHNSERVALGLYILERTHREUSERPROVAL	162
Db	202	TCAAGATACCTCGGCGAACCAATATGGAGCCAAATGCTCTGTGGATTCACATTGACCCAGT	261
QY	163	GLINALAGINALASNALALASUALALALALAGLNGLYLEUALASPLYVALSERPHE	182
Db	262	CAAGCCGAGAGGAAATGCGCTTCGGAGCAGCAAGGGGTTCGGCAAGCGCTTCTTTC	321
QY	183	GLYVALIALASPHALALEUGLNGLINPROPHESERASPSGLYGLNPHESAPLEUALTRPSSER	202
Db	322	CAGTTCCTCATGCTCTGGAGCAACCAATYTCCTGARGGGCAGATTGATCTTGCCTGCTCT	381
QY	203	METGLISERGLYIUNHISMETPROASPLYALALYSPHEVALGLYLIULIULUALAARGVAL	222
Db	382	ATGAGAGTGGTAGCACAATGCCCAACCAACAGACAACTTTGTATAGCGAGCTGGCAGCCGTC	441

OY	223	AlaAlaProGlyAlaIleIleIleIleValThrTrpCysHisArgAspLeuGlyProAsp	242
Db	442	GCAGCTCCAGGAGCAACATATCATATCGAGCTAGTGGCATAGGAACCTGGCCCATCG	501
OY	243	GluGlnSerLeuHisProTrpGluGlnAspLeuLeuLysLysIleCysAspAlaTrpTyr	262
Db	502	GAGGAGCTACCTGAACACCTGACGAGCTGAAATCTTTGAAAAAGATTTTGTATCATATATAC	561
OY	263	LeuProAlaTrpCysSerThrSerAspArgValLysIleuLeuGlnSerLeuSerLeuGln	282
Db	562	CTCCCGGATTTGGTGGTCTCTCCCTCGGATTAATGCAAGATTTGGCGAGCTCATTTGCTTTGAG	621
OY	283	AspIleLysSerGluAspTrpSerArgPheValAlaIleProPheTrpProAlaIleValIleArg	302
Db	622	CATATCAAAACAGCGCGAGCTGGTGTGAAAACGACGCGCCGTTCTGGCTCTGTCTCATCCAA	681
OY	303	SerAlaPheThrTrpLysGly	309
Db	682	TGAGCACTGACACTGGAAAAGCC	702

RESULT 11	
BG320098/c	
LOCUS	831 bp
DEFINITION	mRNA
	linear
	EST 27-FEB-2001
	Zm03.01d11.A Zm03.AAFC.ECORA.cold-stressed_maize-seedlings Zea mays
ACCESSION	BG320098
VERSION	BG320098.1
KEYWORDS	GI:13149776
SOURCE	Est.
	Zea mays.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (Pages 1 to 83)	Sammonds, J. A., Singh, J. A., Piche, C., Cass, L., 'A', Harris, L. J., Hattori, J. I., Ouellet, T., and Tinker, N. A.	Expressed Sequence Tags from cold-stressed Maize Seedlings Grown Under High Light Intensity	Unpublished (2001)	Contact: Singh, J. A.

FEATURES	Location/Qualifiers
source	1. .831

BASE COUNT	166 a	237 c	213 g	214 t	1	Others
ORIGIN						
Alignment Scores:						
Pred. No.:	1,02e-84				831	
Score:	922.00				Matches: 174	
Percent Similarity:	84.19%				Conservative: 23	
Best Local Similarity:	74.36%				Mismatches: 37	
Query Match:	50.38%				Indels: 0	
DB:	12				Gaps: 0	

US-09-857-613a-28 (1-350) x BG320098 (1-831)

QY 116 LeuArgphea1aSerValSerLincLuarGserLysTrpProLysSerL1eValaSpAl 135
 DB 827 CTTCCGCGGTTCCACGCCCTCAGATGATCCAGAGACACCAAAACATAGTCGATGTC 768
 QY 136 GlyCysGly11eGlyLysSerArqTyLLeuAlaLysLysPheGlyAlaThrSerVal 155
 DB 767 GGTTCGGATGCTGGTGTGATGCTCAAGTACTGCGCAAGAAATACGAGCGCAGTGCACCT 708
 QY 156 Gly1LeuThrLeuSerProValGlnAlaGlnArgAlaAsnAlaLeuAlaAlaGlnGly 175
 DB 707 GGGATCAGCTGGAGCCCTGTTCAAGCCGAGAGGAAATGCTCTCGCTGCGAGCGCAGGG 648
 QY 176 LeuAlaAspLysValSerPheGlnValAlaAspAlaLeuGlnGlnProPheSerAspGly 195
 DB 647 TTTCGGATGAGTACTGCTGCAAGTGTGATGCTCTGAGAGCAACGGTTCTCCGACGGG 588
 QY 196 GlnPheAspLeuValTrpSerMetGlnSerGlyLuh1sMetProAspLysAlaLysPhe 215
 DB 587 CAGTTCGATCTGGTGTGCTGCTCATGAGAGTGGCAGACATGCGCAGCAAGAGAAAGTTT 528
 QY 216 ValGlyLincLeuAlaArgValAlaAlaProGlyAlaL1eL1eL1eL1eL1eL1eL1e 235
 DB 527 GTTAGTGAAGTACAGACGCGCTGCGCTCTGAGAGGACAAATATCATGTCATGTCGTC 468
 QY 236 HisArgAspLeuGlyProAspGlnGlnSerLeuHisProTrpGlnGlnAspLeuLys 255
 DB 467 CATAGAACCTGGATCCATCCGAAACCTGCTAAAGCCCGATGAACTGAGCCTCTGAGG 408
 QY 256 Lys1LeCysAspAlaTrpTyLLeuProAlaTrpCysSerThrSerAspTyLValLysLeu 275
 DB 407 AGCATGTCAGACGCTACTACTCCCGAGCTGTGCTCAGCTTCAAGACTATGTGAACATT 348
 QY 276 LeuGlnSerLeuSerLeuGlnAspLysSerGlnAspTrpSerArqPheAlaLapRo 295
 DB 347 GCCAAGTACTGCTCTGAGATATCAAGACAGCTGCTGTCGAGAGACGTCGCGCCG 288
 QY 296 PheTrpProAlaVal11eArgSerAlaPheThrTrpLysGlyLeuSerSerLeuSer 315
 DB 287 TTTTGGCCCCCTGATAAATATCAGCGCTTACATGGAAGGCTTCACTCTGCTGAGG 228
 QY 316 SerGlyGlnLysThrLysGlyAlaLeuAlaMetProLeuMet11eGlnGlyTyLys 335
 DB 227 ACCGATGGAAGACGATCAGAGCGGATGATGCCCTAATGATCAGCGGCTACAG 168
 QY 336 LysAspLeuLeuLysPheAlaL1eL1eL1eL1eL1eL1eL1eL1eL1eL1eL1eL1e 349
 DB 167 AAGGGGCTCATCAATTCACCATCATCAGCTGTCGCAAGCCT 126
 RESULT 12
 BJ463750 664 bp mRNA linear EST 23-MAY-2002
 LOCUS BJ463750 K. Sato unpublished cDNA library, cv. Haruna Nijo
 DEFINITION germination shoots Hardeum vulgare subsp. vulgare cDNA clone
 bags30f10 5', mRNA sequence.
 ACCESSION BJ463750
 VERSION BJ463750.1 GI:21142257
 KEYWORDS
 SOURCE Hardeum vulgare subsp. vulgare.
 ORGANISM Hardeum vulgare subsp. vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Hordeum.
 REFERENCE 1 (bases 1 to 664)
 AUTHORS Sato, K., Saito, D. and Takeda, K.
 TITLE Barley EST sequencing project in NIG and Okayama Univ
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadasi Shin-i
 Center for Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856

FEATURES
 source
 1. 664
 /organism="Hardeum vulgare subsp. vulgare"
 /cultivar="Haruna Nijo"
 /db_xref="taxon:112509"
 /clone_lib="K. Sato unpublished cDNA library, cv. Haruna
 Nijo germination shoots"
 /tissue_type="shoots"
 /dev_stage="germination"

BASE COUNT 158 a 177 c 186 g 143 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1,8e-82 Length: 664
 Score: 899.00 Matches: 168
 Percent Similarity: 86.24% Conservative: 20
 Best Local Similarity: 77.06% Mismatches: 30
 Query Match: 49.13% Indels: 0
 DB: 13 Gaps: 0

US-09-857-613a-28 (1-350) x BJ463750 (1-664)

QY 73 TyrAspGlnSerSerGlyLeuTrpGlnAsnTrpGlyAspHisMetHisGlyPhe 92
 DB 4 TACAGCAGTCCGCTCCGCGCTGTGGAGAGACATCTGGCGCAGACATGCACACCGCTTC 63

QY 93 TyrAspSerAspSerThrValSerLeuSerAspHisArgAlaAlaGln11eArgMet11e 112
 DB 64 TACGACTGTGGCGAGCGCCCTCCATGTCCGACACCGCGCGCCAGATCCGATGATC 123

QY 113 GlnGlnSerLeuArgPheAlaSerValSerGlnGlnArgSerLysTrpProLysSer1e 132
 DB 124 GAGAGGCCCTGCTTCCCGCCGCGCCGAGACGCCCAACCAACCAACCAATTT 183

QY 133 ValAspValGlyCysGlyL1eGlyLysSerArqTyLLeuAlaLysLysPheGlyAla 152
 DB 184 GTTGATGTTGATGCGCAATCGCGGTGATCTCAAGTACTGCGCAAAATATGAGCA 243

QY 153 ThrSerValGlyL1eThrLeuSerProValGlnAlaGlnArgAlaAsnAlaLeuAla 172
 DB 244 CAGTCTCTGGATCACATTCAGCCAGCTGCAAGCGAGAGAGAAATGCCCTTACGCGG 303

QY 173 AlaglnGlyLeuAlaAspLysValSerPheGlnValAlaAspAlaLeuGlnGlnProPhe 192
 DB 304 GCACAGGGGTTGGCAGACAGAGGCTTCTTCCAAAGTCTGATGCTGAGACACCAATT 363

QY 193 SerAspGlnPheAspLeuValTrpSerMetGlnSerGlyLuh1sMetProAspLys 212
 DB 364 CCTGATGGCAGTTTGATCTTGTGCTGTGAGAGAGTGTGAGACATGCCAACA 423

QY 213 AlAlasPheValGlyGlnLeuAlaArgValAlaAlaProGlyAlaL1eL1eL1eL1eL1eL1e 232
 DB 424 CAGAGTGTGAGTGAAGTCTGACGCGCTCCAGCTCCAGGAGCAACATATCATCTCG 483

QY 233 ThrTrpCysHisArgAspLeuValTrpSerMetGlnSerGlyLuh1sMetProAspLys 252
 DB 484 ACCGTGTCATGACACCTTCGCTATCTGAGACTCATCTGAAACCTGACAGCTGAT 543

QY 253 LeuLeuLysLys11eCysAspAlaTrpTyLLeuProAlaTrpCysSerThrSerAspTyL 272
 DB 544 CTTTGAAGAAATGTTGATCATATACCTCCGAGATTTGGTGGCTCGCGCTGATTT 603

QY 273 ValLysLeuLeuGlnSerLeuGlnAspLysLysSerGlnAspTrpSer 290
 DB 604 GTCAAGATTGCCAGTCATGCTCTTGGATATCAAAAACGGCTGACTGCT 657

RESULT 13
 BJ463750 515 bp mRNA linear EST 07-MAY-2002
 LOCUS BJ463750 K. Sato unpublished cDNA library, cv. Haruna Nijo
 DEFINITION saoz3e12.y1 gm-cl054 Glycine max cDNA clone SOYBEAN CLONE ID:

Gm-c1054-5207 5' similar to TR:Q9XIP9 Q9XIP9 GAMMA-TOCOPHEROL
METHYLTRANSFERASE. [1] , mRNA sequence.

ACCESSION BQ273302
VERSION BQ273302.1 GI:20498372
KEYWORDS EST.
SOURCE soybean.

ORGANISM Glycine max

Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
Rosidae: eurosids I: Fabales; Fabaceae: Papilionoideae; Phaseoleae:
Glycine.

REFERENCE 1 (bases 1 to 515)

AUTHORS

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corryell, V., Khanna
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers,
Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk,
R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
R., Waterston, R. and Wilson, R.
Public Soybean EST Project

TITLE Unpublished (1999)
JOURNAL Contact: Shoemaker R/Public Soybean EST Project
COMMENT Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Resgen, Invitrogen Corp. 2130
South Memorial Parkway Huntville, AL 35801 For further information
call: (800)-533-4363 or contact: cculeresgen.com web site:
www.resgen.com
Seq primer: -40RP from Gibco
High quality sequence stop: 421.

FEATURES
source Location/Qualifiers

1..515

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="SOYBEAN CLONE ID: Gm-c1054-5207"

/clone_11b="Gm-c1054"

/tissue_type="leaf, 3 week old, greenhouse grown"

/lab_host="DH10B"

/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
XhoI; The Harosoy NIL was constructed and seed was
provided by Dr. J. Specht, University of Nebraska
(Shoemaker and Specht, 1995). The cDNA library was
constructed from mRNA isolated from leaf tissue at various
developmental stages of 3 week old greenhouse grown
plants. Complementary DNA was synthesized from mRNA using
a primer consisting of a poly(dT) sequence with a XhoI
restriction site and a 3' anchor. EcoRI adapters were
ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into DH10B host cells (GibcoBRL). This
library was constructed in cooperation with Dr. Paul
Keim's laboratory at Northern Arizona University."

BASE COUNT 124 a 113 c 135 g 143 t

ORIGIN

Alignment Scores:

Pred. NO: 4.56e-81

Score: 884.00

Percent Similarity: 99.42%

Best Local Similarity: 99.42%

Query Match: 48.31%

Length: 515
Matches: 170
Conservative: 0
Mismatch: 1
Indels: 0
Gaps: 0

US-09-857-613a-28 (1-350) x BQ273302 (1-515)

OY 89 HSHISGLYPhETyRASPSeRAsPSeRthValSeRSeRAsPShISArgAlaIaGln 108
|||||
Db 3 CACCATGGCTTTTAAAGATCGGATTCACACTGTTTGGCTTGGATCATTGCTGCTCAG 62

OY 109 IleaRgMetIleGlnGlnSeRSeRLeuArGpheaAlaSeRValSeRGlunArgSeRlyStrp 128
|||||
Db 63 ATCCGAATAGATCCAGAGCTCTCTCCCTTCCCTCTGTTTCTAGACGCTACTAATAG 122
OY 129 ProLySeRlIeValaSpValGlycSGlylIeglyGlySeRSeRArgTyRLeuAlaLyS 148
|||||
Db 123 CCCAAGAGTATAGTTGATGTTGGGTGTGCATAGTGCGCAGCTAGTACCTGGCCAG 182
OY 149 LysPhGlyAlaIhSeRSeRValGlylIeThrLeuSeRProValGlnAlaGlnArgAlaAsn 168
|||||
Db 183 AATTGGAGCAACGAGCTAGCATCCTGAGTCCGTTCAAGCTCAAGAGCAAAAT 242
OY 169 AlAlaValAlaIaIaGlnGlnLeuAlaAspLySValSeRPhedGlnValAlaAspAlaVal 188
|||||
Db 243 GCCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 302
OY 189 GlnGlnProPhSeRAspGlyGlnPheaSpLeuValTrpSeRMetGlnSeRGlunHis 208
|||||
Db 303 CAGCAACCATTTCTCTGAGCGGCGAGTTGATCTGTGCTGCTGCTGCTGCTGCTGCTGCT 362
OY 209 MetProAspLyAlaLySPhSeRValGlylGlnLeuAlaArgValAlaAlaProGlyAlaIle 228
|||||
Db 363 ATGCTGCAAGCAAGCTAAGTTGTTGGAGAGTTAGCTGGGAGCAGCAGCAGTGCCTACT 422
OY 229 IleaIleIleValIhTrPpCShISArGAsPLeuGlyProAspGlnGlnSeRLeuHisPro 248
|||||
Db 423 ATAAATATATGTAACATGATGTCACAGAGATCTTGGCCCTGACGACACATCTTACTCA 482
OY 249 TrpGlnGlnAspLeuLeuLySlylIecyAsp 259
|||||
Db 483 TGGAGCAAGATCTCTTAAGAAGATTTCGAT 515

RESULT 14

LOCUS

DEFINITION

VERSION

KEYWORDS

SOURCE

ORGANISM

Glycine max

soybean.

REFERENCE

AUTHORS

1 (bases 1 to 532)

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corryell, V., Khanna

A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,

Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers,

Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk,

R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann

R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Resgen, Invitrogen Corp. 2130
South Memorial Parkway Huntville, AL 35801 For further information
call: (800)-533-4363 or contact via email: cculeresgen.com
Insert Length: 1413 Std Error: 0.00
High quality sequence stop: 429.

TITLE

JOURNAL

COMMENT

FEATURES
source Location/Qualifiers

1..532

/organism="Glycine max"

/db_xref="taxon:3847"

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 19, 2003, 05:12:15 ; Search time 2336 Seconds

(without alignments)
4360.438 Million cell updates/sec

Title: US-09-857-613a-28

Perfect score: 1830
Sequence: 1 MATVRIPTISCIHIFRS.....IEGYKDLIKFAITCRKPE 350

Scoring table:

BLOSUM62
Xgapop 10.0 , Ygapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL-frame_p2n -DEV-xmlh
-O-/cgn2.1/USPTO/US09857613/runat_12032003_100036_6013/app_query.fasta_1.519
-DB-GenBml -QFMT-fastap -SUFFIX-rige -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=45
-UNITS-bits -START=1 -END=1 -MATRIX-biosum62 -TRANS-human40.cdi -List=45
-OCALIGN=200 -THR.SCORER=1 -THR.MAX=100 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09857613 -ECGN_1_1_1687 -etunal_12032003_100036_6013 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAPP -LARGEQUERY -NEG.SCORER=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6 -FCGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenBml:.*
1: gb_da.*
2: gb_hlg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_da.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vl.*
30: em_hlg_hum.*
31: em_hlg_inv.*
32: em_hlg_other.*
33: em_hlg_mus.*
34: em_hlg_pln.*
35: em_hlg_rtd.*
36: em_hlg_mam.*
37: em_hlg_vrt.*
38: em_sy.*
39: em_hlgo_hum.*
40: em_hlgo_mus.*
41: em_hlgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1159	63.3	1369	AF213481	AF213481 perilla f
2	1154.5	63.1	1047	AY090280	AY090280 Arabidops
3	1154.5	63.1	1303	AY049258	AY049258 Arabidops
4	1151.5	62.9	1350	AX417728	AX417728 Sequence
5	1151.5	62.9	1350	AF104220	AF104220 Arabidops
6	1147.5	62.7	1296	AY087138	AY087138 Arabidops
7	981.5	53.6	11835	AC006193	AC006193 Arabidops
8	758	41.4	125422	AP003994	AP003994 Oryza sat
9	659	36.0	343550	AP003587	AP003587 Nostoc sp
10	647.5	35.4	134199	SYCSLRF	SYCSLRF Nostoc sp
11	579	31.6	108553	AF527809	AF527809 Sorghum b
12	416.5	22.8	298750	AP005375	AP005375 Neurospyn
13	392.5	21.4	334520	AP003588	AP003588 Nostoc sp
14	392	21.4	930	AX073657	AX073657 Sequence
15	392	21.4	957	AX073657	AX073657 Sequence
16	392	21.4	957	AX417730	AX417730 Sequence
17	392	21.4	974	AX073659	AX073659 Sequence
18	392	21.4	145709	D90914	D90914 Synchocyst
19	294	16.1	25681	SAE414559	SAE414559 Saccharot
20	287	15.7	63734	AF127374	AF127374 Streptomyl
21	282.5	15.4	90445	AF040570	AF040570 Amycolato
22	270	14.8	12637	AF323753	AF323753 Streptomyl
23	268.5	14.7	125561	NC966	AL513463 Neurospor
24	261	14.3	2122	A60301	A60301 Sequence 1
25	261	14.3	2122	AR144762	AR144762 Sequence
26	247	13.5	1276	NT071107	NT071107 Nicotiana t
27	244.5	13.4	5998	AB032524	AB032524 Streptomyl
28	242	13.2	30000	AX250261	AX250261 Sequence
29	241	13.2	1392	TAU60754	U60754 Trifolium ae
30	237	13.0	1410	E07846	E07846 DNA sequenc
31	234.5	12.8	349498	AP003002	AP003002 Mesorhizo
32	232.5	12.7	1698	AF216283	AF216283 Acclimopol
33	231	12.6	1629	AF042332	AF042332 Oryza sat
34	228	12.5	11714	AE012928	AE012928 Chlorobiu
35	227	12.4	1383	ZMU79669	ZMU79669 Zea mays en
36	227	12.4	40065	SPBC16E9	SPBC16E9 S.pombe chr
37	226	12.3	840	AF216282	AF216282 Halorhodo
38	226	12.3	1080	AY113031	AY113031 Arabidops
39	226	12.3	1302	AF375397	AF375397 Arabidops
40	226	12.3	1497	AF045570	AF045570 Zea mays
41	226	12.3	97554	AC009978	AC009978 Genomic s
42	225	12.3	6085	A60304	A60304 Sequence 4
43	225	12.3	6085	AR144763	AR144763 Sequence
44	224	12.2	1094	D89131	D89131 Schizosacch
45	223.5	12.2	1267	NT071108	NT071108 Nicotiana t

RESULT 1

ALIGNMENTS

AF213481 1369 bp mRNA linear PLN 02-DEC-2001
 DEFINITION Perilla frutescens gamma-tocopherol methyltransferase (TMT) mRNA,
 complete cds.
 ACCESSION AF213481
 VERSION AF213481.1 GI:17224291
 KEYWORDS
 SOURCE Perilla frutescens.
 ORGANISM Perilla frutescens
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Lamiales; Lamnaceae; Perilla.
 1 (bases 1 to 1369)
 Kim, K.-H., Hwang, S.-K. and Hwang, Y.-S.
 Cloning of Perilla gamma-tocopherol methyltransferase
 Unpublished
 2 (bases 1 to 1369)
 Kim, K.-H., Hwang, S.-K. and Hwang, Y.-S.
 Direct Submission
 Submitted (09-DEC-1999) Division of Biochemistry, National
 Institute of Agricultural Science and Technology, 249 Seodun-dong,
 Wonsun-gu, Suwon, Kyunggi-do 441-707, Republic of Korea
 FEATURES
 source
 1..1369
 /organism="Perilla frutescens"
 /cultivar="Okdong"
 /db_xref="taxon:48386"
 /tissue_type="developing seed"
 1..1369
 /gene="TMT"
 89..1198
 /gene="TMT"
 /codon_start=1
 /product="gamma-tocopherol methyltransferase"
 /protein_id="AAL36933.1"
 /db_xref="GI:17224292"
 /translation="MAEATPGICCTGMRGCVAAFTYISIKPATALLVGTTRKTS
 ITSESTLSRTGRARAPPTMSLAAAEEMETEMETLRKIAFEYESSGVWNINGMDH
 MHNGYEPADVISIDHRAAQIRMEESIRFASFSTTEPKRNIDYDVGCGIGSSR
 VLARKYGAKLRAITLSPVQAQRAQOQLADAGLNGKVFEEVADALNOPEPGKFDLY
 MSYEGEHMPDKRFVNEELRVAAFGRIITVWCHROLSPSEESIROEKOLLNKIC
 SAYVLPAMSTADYVKLLDLSIMEDIKSDMSDHPAIPPAVYKSLTKTKITLSLRS
 CMKTRGAMVHMLMEIGYKVKYFAITTRKRPAS"
 BASE COUNT 380 a 306 c 332 g 351 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2 39e-92 Length: 1369
 Score: 1159.00 Matches: 233
 Percent Similarity: 78.11% Conservative: 31
 Best Local Similarity: 68.93% Mismatches: 56
 Query Match: 63.33% Indels: 18
 DB: 8 Gaps: 5
 US-09-857-613a-28 (1-350) x AF213481 (1-1369)
 QY 15 ILEHSTHPhEARySerInSerProARyTHrPhEAAlARgIleArYValGIyProARy 34
 DB 221 ATTACTTCTTTTCCACACACACCTCCACAGACACGTGGAGACACACGT-----CGC 271
 QY 35 SerTPAlARProIleARgAlASerAlAlASerSerGIuARgGIyGIuIleValLeuGIu 54
 DB 272 CCGAGCATGAGCCTGAAACCCCTCGCGGAGATGAG----- 310
 QY 55 GIuIySProlYLySAspLySylSylSylLeuGIuIyGIyIleAlGIuPhETyAsp 74
 DB 311 -----ACGAGATGAGGAGACCTTGCGTAAGGAGATTCGGAGTTCTACGAC 355
 QY 75 GluSerSerGIyLeuTPolUAsnIleTPGIYAspHISMetHIShtSGIyPhETyAsp 94
 DB 356 GAGTCGTGGGGGTGGGAGAAACATATGGGAGACCAATGCACACGCGCTTTTACGAG 415
 QY 95 SerAspSerTHrValSerLeuSerAspHISARgAlAlAGInIleARgMetIleGIuIn 114

DB 416 CCGCGCGCGACGTCCTCATCTCCGACACATCGCGCGCGCGCGCATTCATTCAGAG 475
 QY 115 SerLeuARgPhEAAlASerValSerGIu---GIuARgSerLySTPProLySerIleVal 133
 DB 476 TCCCTCCGATTCGCTCTCTCTCCGATTAACATACAGAGAAACCGAATAATTAAGT 535
 QY 134 AspValGIySGIyIleGIyIleGIySerSerARgTHrLeuAlAlLySylPhGIyAlA--- 152
 DB 536 GATTCGAGATGTGTATAGAGGAGGATTCATGATTCGGAGAAATAATAGGGCTTAA 595
 QY 153 ThrSerValGIyIleTHrLeu---SerProValGIuAlAGrAlAlASnAlLeuAla 171
 DB 596 TTGCTAGGGGTATATCTCTTCACAGCCCTCTCGAAGCGCAGAGACCTCAACAGCTTCT 655
 QY 172 AlAlAGInGIyLeuAlAlAspLyValSerPhGIuValAlAlAspAlaLeuGIuInPro 191
 DB 656 GATGCTCAAGCATTAATAGGAGCATTCCTTGAAGTGTGATGCTTGAACCAACCA 715
 QY 192 PhSerAspGIyGIuPhEAAspLeuValTrpSerMetGIuSerGIyGIuHISMetProAsp 211
 DB 716 TTTCCTGAAGGAGATTTGATGTGTTGTGATGAGAGATGGAGAACACATCTCGAT 775
 QY 212 LysAlAlAspPhEAAlGIyGIuLeuAlArGIyAlAlAlAProGIyAlAlIleIleIle 231
 DB 776 AAGAAAGATTTGTAATAGAGCTGTGCTGTGCTGTCTGTGTGAGAAATATATCTC 835
 QY 232 ValIThrTPCySHISARgAspLeuGIyProAspGIuInSerLeuHISProTPGIuIn 251
 DB 836 GTTACATGATGCGCACAGGAGCTATCATCTGTAAGAAATCTTTCGCAAGAGAGAA 895
 QY 252 AspLeuLeuLySylIleCyAspAlATyTrpLeuProAlArTPCySerTHrSerAsp 271
 DB 896 GATTTCGTAACAAATATATAGCTTATATCTTCCAGCATGCTGTCTACGCTGAC 955
 QY 272 TyrValLySylLeuGIuInSerLeuSerLeuGIuInAspIleLySylSerGIuAspTrpSerArg 291
 DB 956 TATGTCAAATTAATACGACCTCTCTCAATGAGAGACATTAAGTCTGCAGACTGTCTGAC 1015
 QY 292 PheValAlArProberTPProAlAlValIleArSerAlArberTHrTPLySGIyLeuSer 311
 DB 1016 CAGTCGCTCATTTTGGCGCGGAGATTAAGCGGCAATGATGAGAGCGGATTAAC 1075
 QY 312 SerLeuLeuSerSerGIyGIuIyTHrIleLySGIyAlAlAlMetProLeuMetIle 331
 DB 1076 TCACCTGTAGAGAGCGGATGGAAGACTATTAAGAGACCAATGTGATCATTCATGATTC 1135
 QY 332 GluGIyTyIleLySylAspLeuIleLySylPheAlAlIleIleTHrCySArgLySylPro 349
 DB 1136 GAAGGATATAGAAAGGCGGTGATCAATTTGCCATCTTATCATGCCGAAACCT 1189
 RESULT 2
 AY090280 1047 bp mRNA linear PLN 24-MAR-2002
 LOCUS Arabidopsis thaliana Atlg64970/F13011_27 mRNA, complete cds.
 ACCESSION AY090280
 VERSION AY090280.1 GI:19699149
 KEYWORDS FLI_CDNA.
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 1 (bases 1 to 1047)
 Kim, C.-J., Chen, H., Cheuk, R., Meyers, M.-C., Shinn, P., Banh, J.,
 Bowser, L., Carninci, P., Chang, E., Dale, J. M., Goldsmith, A. D.,
 Hayashizaki, Y., Ishida, J., Jones, T., Kamuya, A., Karlin-Neumann, G.,
 Kawai, J., Lam, B., Lee, J. M., Lin, J., Miranda, M., Narusaka, M.,
 Nguyen, M., Onodera, C. S., Palm, C. J., Quach, H. L., Sakurai, T.,
 Satou, M., Seki, M., Southwick, A., Tang, C. C., Toriumi, M., Wu, H. C.,
 Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinzaki, K., Davis, R. W.,
 Theologis, A. and Ecker, J. R.
 Arabidopsis ORF clones
 TITLE

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 1047)

Kim, C.J., Chen, H., Cheuk, R., Meyers, M.C., Shinn, P., Banh, J.,
Bower, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D.,
Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Katlin-Neumann, G.,
Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M.,
Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T.,
Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C.,
Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinzaki, K., Davis, R.M.,
Theologis, A. and Ecker, J.R.

TITLE
JOURNAL

Direct Submission
Submitted (14-Mar-2002) Salk Institute Genomic Analysis Laboratory
(Signal), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA

COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the
sequencing and annotation of the RFL CDNA. (RFL CDNA: Kim, C.J., Chen, H.,
Cheuk, R., Meyers, M.C., Shinn, P., Banh, J., Bower, L., Chang, E.,
Dale, J.M., Goldsmith, A.D., Jones, T., Katlin-Neumann, G., Lam, B.,
Lee, J.M., Lin, J., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J.,
Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C.,
Yamada, K., Yamamura, Y., Yu, G., Yu, S., Davis, R.M., Theologis, A.,
and Ecker, J.R.

The Salk, Stanford, PGSC (SSP) Consortium members carried out the
sequencing and annotation of the RFL CDNA. (RFL CDNA: Kim, C.J., Chen, H.,
Cheuk, R., Meyers, M.C., Shinn, P., Banh, J., Bower, L., Chang, E.,
Dale, J.M., Goldsmith, A.D., Jones, T., Katlin-Neumann, G., Lam, B.,
Lee, J.M., Lin, J., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J.,
Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C.,
Yamada, K., Yamamura, Y., Yu, G., Yu, S., Davis, R.M., Theologis, A.,
and Ecker, J.R.

FEATURES
source

CDS

1..1047
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
/clone="U13225"
/note="This clone is in pUNI 51
ecotype: Columbia"
1..1047
/note="gamma-tocopherol methyltransferase"
/codon_start=1
/product="At1g64970/F13011_27"
/protein_id="At1g64970.1"
/db_xref="GI:19699150"
/translation="MKATLADESLTSLPYRTNSFGSKSLFRSPSSSSVSMTT
RGNVAATAATSTELAKGIAEFYNETSGIWEIWDHMHGFYDSDSVOLDSGHK
EAOIRMIESTLRAGVTDEEEKIRKVVVDCGSGSRYSKSGAECIGITLSPV
OAKRANDLAASOLAKHASFQVADALDOPEDCKFLVSMESGEMHPDKAFVKELY
RYAAPGCRITITWCKRNLNAGEALQDPONLIDICTFTYIPACSCDDYNNILDS
HSLDITKADMSBNVAFPMFAYTRTALTKGVLVSLRSGKSTKGLTNPMLTEGTRK
GVTKFGITRQKPL"

BASE COUNT
ORIGIN

267 a 223 c 269 g 288 t
Alignment Scores:
Pred. No.: 4.22e-92 Length: 1047
Score: 1154.50 Matches: 231
Percent Similarity: 75.43% Conservative: 30
Best Local Similarity: 66.76% Mismatches: 65
Query Match: 63.09% Indels: 21
DB: 8 Gaps: 4

US-09-857-613a-28 (1-350) x AY090280 (1-1047)

OY 8 ProthrlleScysIleHlsIleThrpheargSerInserProargThrpheala 27
DB 53 CCAACCTCTTCTTGGGTCAAGCATCGCTTCTTT-CGGTCTCCATCTCTCTCC 111
OY 28 ArgIleargVal---GlyProargSerTrpAlaProIleargAlaSerAlaIleSerSer 46
DB 53 CCAACCTCTTCTTGGGTCAAGCATCGCTTCTTT-CGGTCTCCATCTCTCTCC 111

DB 112 TCAGTCTATGACGACCAACGCGTGAAGAGTGGCTGCTGCTCTCTACATCCACT 171
OY 47 GUAAGGlyGluIleValleuGluGlnLysProLysAspLysLysLeuGln 66
DB 172 GAG-----GGGCTAAGA 183
OY 67 LysGlyIleAlaGluPheTyraSpGluSerSerGlyLeuTrpGluAsnIleTrpGlyAsp 86
DB 184 AAGGAATAGCGGAGTTCATCATGATGAACACTTCGGTTTGGGAAAGATTTGGGGAGAT 243
OY 87 HisMetHisIleGlyPheTyraSpSerSerThpValSerLeuSerAsp-----His 104
DB 244 CATATGCAATCATGCGTTTATGACCTGATCTTCTTCTTCACTTCTGATTTGCTGAC 303
OY 105 ArgAlaGluGlnIleArgMetIleGlnLysLeuArgPheAlaSerValSer---Glu 123
DB 304 AAGCAAGCTAGATCCGATGATGATGAAGACTCTCTGCTTTCGGGTATGATGATGAA 363
OY 124 GluArgSerLysTrpProLysSerIleValAspValGlyCysGlyIleGlySerSer 143
DB 364 GAGGAGAGCAAAAGATTAAGAAAGTATGATGATGATGATGATGATGATGATGATGAT 423
OY 144 ArgTyrlleuAlaLysPheGlyAlaThrSerValGlyIleThrLeuSerProValGln 163
DB 424 AGATATCTTCTCTTAATTTGAGCTGAATGATGATGATGATGATGATGATGATGATGAT 483
OY 164 AlaGlnArgAlaAsnAlaLeuAlaAlaAlaGlnGlyLeuAlaAspLysValSerPheGln 183
DB 484 GCCAAGAGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 543
OY 184 ValAlaAspAlaLeuGlnGlnIleProPheSerAspGlyGlnPheAspLeuValTrpSerMet 203
DB 544 GTTGGGATGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 603
OY 204 GluSerGlyLysMetProAspLysAlaLysPheValGlyLeuAlaArgValAla 223
DB 604 GAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 663
OY 224 AlaProGlyAlaIleIleIleIleValThrTrpCysHisArgAspLeuGlyProAspGlu 243
DB 664 GCTCAGAGAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 723
OY 244 GlnSerLeuHisProTrpGluGlnAspLeuLysLysIleCysAspAlaTyrlleu 263
DB 724 GAACCTTTCGACGCTGCGGAGCAAAACATCTTGCACAAATCTTACACCTTCTATCTC 783
OY 264 ProAlaTrpCysSerThrTrpSerAspTrpValLysLeuGlnIleSerLeuGlnAsp 283
DB 784 CCGGCTTGGTCCACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 843
OY 284 IleLysSerGluAspTrpSerArgPheValAlaProPheTrpProAlaValIleArgSer 303
DB 844 ATTAAGTGTGCGGATTTGTCAGAGAAGTACCTCTCTTCTGCTGCGGATTTACGAGCT 903
OY 304 AlaPheThrTrpLysGlyLeuSerSerLeuLeuSerSerGlyGlnLysThrIleLysGly 323
DB 904 GCATTATACATGAGAGGCGCTTGTCTCTGCTGCTGATGATGATGATGATGATGATGAT 963
OY 324 AlaLeuAlaMetProLeuMetIleGluGlyTyrlleLysAspLeuIleLysPheAlaIle 343
DB 964 GCATTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1023
OY 344 IleThrCysArgLysPro 349
DB 1024 ATCACTTGCAGAGCCA 1041

RESULT 3
AY049258 1303 bp mRNA linear PLN 09-AUG-2001
LOCUS AY049258
DEFINITION Arabidopsis thaliana At1g64970/F13011_27 mRNA, complete cds.
ACCESSION AY049258
VERSION AY049258.1 GI:15146233
KEYWORDS F13011_27

QY	324	AltaLualamerProleummetillegluclylryrlylsyaspLeuilellyleaphleatle	343
Db	1031	GCATTGACATCCATTGATGATTGAAGGTTTCAAGAAAGGTGTCATTAAAGTTTGGTATC	1090
QY	344	lIethrCysArgLysPro	349
Db	1091	ATCACTTGCCAGACCA	1108
RESULT 4			
AA417728			
LOCUS	AX417728	1350 bp	DNA
DEFINITION	Sequence 19 from Patent WO0231173.		
ACCESSION	AX417728		
VERSION	AX417728.1	GI:21522874	
KEYWORDS			
SOURCE			
ORGANISM			
	thale cress.		
	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidops.		
REFERENCE			
AUTHORS	1 Ebneeth, M., Kunze, I. and Geiger, M.		
TITLE	Improved method for the biosynthesis of vitamin E		
JOURNAL	Patent: WO 0231173-A 19 18-APR-2002;		
FEATURES			
Source	Location/Qualifiers		
	1..1350		
	/organism="Arabidopsis thaliana"		
	/db_xref="taxon:3702"		
	63..1109		
	/note="coding for gamma-tocopherol methyltransferase"		
	/codon_start=1		
	/protein_id="CAD35425.1"		
	/db_xref="GI:21522875"		
	/translation="MKATLAAPSSLTSLPRTNSSFGSKSLFRBSSSSVSMTT		
	RGNVVAATAATSTALRKGIAGFYNETSGLWEIEGMHGFYDPDSVOLSDGHR		
	EAQIVMEESLRFAGVTDEEERKIKIKVDVCGGSSRYLAKFGECIGITLSPV		
	QAKRNDLAAASLHKASFOYADLADPFDGKFDLWMSGEHMDKAFYKELY		
	RYAAGKRIITWCHNRLLSGEFAIQLQWEDNIIDKICTPYLPMCSDDPYVULQS		
	HSLODIKCADMSENVAPFPWPAYIRALTLWKGIVSLRSGMSIKGALTMLMIEGKK		
	GVTKGIITCKPL"		
BASE COUNT	374 a 278 c 314 g 384 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	1,07e-91	Length:	1350
Score:	1151.50	Matches:	230
Percent Similarity:	75.43%	Conservative:	31
Best Local Similarity:	66.47%	Mismatches:	65
Query Match:	62.92%	Indels:	21
DB:	6	Gaps:	4
US-09-857-613A-28 (1-350) x AX417728 (1-1350)			
QY	8	ProThrIleSerCysIleHisIleHisThPheArgSerGlnSerProArgThrPheAla	27
Db	115	CCACACTCTCTTCGGGCTCAAGTCAATCGCTTCTTT-CGGCTTCACATCCCTCCTCC	173
QY	28	ArgIleArgVal--GlyProArgSerTrpAlaProIleArgAlaSerAlaSerSer	46
Db	174	TCAGTCTCATGACGACACACGCGTGAACAGTGGTGGCGGCTCTCTCATTCACACT	233
QY	47	GluArgGlyCyluIleValLeuGluGlnGlyProLysLysAspAspLysLysLysLeuGln	66
Db	234	GAG-----	245
QY	67	LysGlyIleAlaGluPheThyArgAspGlnSerSerGlyLeuTrpGluAsnIleTrpGlyAsp	86
Db	246	AAAGGATATACCGGAGTCTCTCATGAAACTTTCGGGTTTGGGGAACAGATTGGGAGAT	305
QY	87	HisMetHisIscLysIlePheThyArgSerAspSerThValSerLeuSerAsp-----His	104

Db	306	CANATGCAATCATGGCTTTTATGACCCCTGATCTCTCTGTCACATCTTGATTCGATTCGAC	365
Qy	105	ArgAlaIaIaGlnIleArgMetIleGlnIleSerLeuArgPheAlaSerValSerGlu	123
Db	366	AAGGAGCGATCGATCCGATGATGAAAGAGCTCTCCCTTCGCGCGGTACTGACGAA	425
Qy	124	GluArgSerLysTrpProLysSerIleValAspValGlyGlyIleGlyGlySer	143
Db	426	GAGGAGGAGAAAGATGAAGAAATAGTGGATGTGGGTGGATGGAGAGCTCA	485
Qy	144	ArgTyrLeuAlaLysLysPheGlyValThrSerValGlyIleThrLeuSerProValGln	163
Db	486	ACATATCTTCCTGCTAAATTTTGAGACCTGAATGACATGGCATTTACTCTCAGCCGTTCAG	545
Qy	164	AlaGlnArgAlaAsnAlaLeuAlaIaIaGlnGlyLeuAlaAspLysValSerPheGln	183
Db	546	GCCAGAGAGGCCAAATGATCTCCGCGCTCTCAATCACTCTCTCAATAGAGCTTCCTCA	605
Qy	184	ValAlaAspAlaLeuGlnGlnProPheSerAspGlyGlnPheAspLeuValTrpSerMet	203
Db	606	GTTGGGGAATGCGTGGATGATGACCATTCGACAGATGGAAATTCGATCTGATGGTGGATG	665
Qy	204	GluSerGlyGlnIleMetProAspLysAlaLysPheValGlyGlyLeuAlaArgValAla	223
Db	666	GAGATGCGTGAACATATGCTGCTACAGGCCAAAGTTTGTAAGAAGCTTGACGTGTGGCG	725
Qy	224	AlaProGlyAlaIleIleIleValThrTrpCysHisArgAspLeuGlyProAspGlu	243
Db	726	GCTCCAGAGGATGATATATATAGTGCATGCTGCCATAGCAAACTGTACGCGGGAGT	785
Qy	244	GlnSerLeuHisProTrpGlnAspLeuLysLysIleCysAspAlaTyrTrpLeu	263
Db	786	GAGAGCTTGGACCCCGGGAGCAAAACATCTTGACAAATCTGTAAAGACGTTCTATCTC	845
Qy	264	ProAlaTrpCysSerThrSerAspTyrValLysLeuGlnSerLeuSerLeuAlaAsp	283
Db	846	CCGGCTGGTGGCTCCACCGCATGATATATGTCACACTTGCTTCATCCCATCTCTCCAGAT	905
Qy	284	IleLysSerGluAspTrpSerArgPheValAlaProPheTrpProAlaValIleArgSer	303
Db	906	ATTAACTGCGGATGGTGCACAGAACGATGCTCTTCTGCGCTGCGTATACGAGCAT	965
Qy	304	AlaPheThrTrpLysGlyLeuSerSerLeuLeuSerSerGlyGlnLysThrIleLysGly	323
Db	966	GCATTACATGGAAAGGCGCTTGCTGCTGCTTCGCTAGCGATGAAAGATTTTAAAGA	1025
Qy	324	AlaLeuAlaMetProLeuMetIleGlnGlyTyrLysAspLeuIleLysPheAlaIle	343
Db	1026	GCAATTACACATCCATGATGATGAAAGGTTACAAAGAAAGGTGATTATAGTTTGATATC	1085
Qy	344	IleThrCysArgLysPro	349
Db	1086	ATTCACCTTGGCAGAGGCCA	1103
RESULT 5	AF104220	1350 bp	linear
LOCUS	AF104220	1350 bp	mRNA
DEFINITION	Arabidopsis thaliana gamma-tocopherol methyltransferase (g-TMT)		
ACCESSION	AF104220		complete cds.
VERSION	AF104220.1		GI:4106537
KEYWORDS			
SOURCE	Arabidopsis thaliana.		
ORGANISM	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids II: Brassicales: Brassicaceae: Arabidopsis.		
REFERENCE	Shintani,D.K. and Dellapenna,D.		
AUTHORS	Elevating Vitamin E content of plants through metabolic engineering		
JOURNAL	Science (1998) In press		
REFERENCE	2 (bases 1 to 1350)		
AUTHORS	Shintani,D.K. and Dellapenna,D.		

C-terminal end of the encoded protein. Please note that these cDNA sequences are derived from the Ms or laer ecotypes and therefore may contain polymorphisms when compared to sequences from Col-0. Genes carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5' sequences, selection of clones, and sequence assembly.

FEATURES

source

1. 1296
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="32143"

CDS

68. 1114
/codon_start=1
/product="gamma-tocopherol methyltransferase"
/protein_id="AA064696.1"
/db_xref="GI:21592747"
/translation="MKATLAAPSLTSLPYRTNSSFSGSKSLIFNSPSSSVSMVTT
KGNVAVAAATSTALRKIAEFYNETSGIMEIWDHMHGFYDPSVSLSDSGHK
EAOIEMIESLRFACVTDDEEEKIKKVVDCGJGSSRSYLSKAGACIGITLSPV
OAKRNDIAAOSLSHKASFOYDALDKPFDKPDLYMSGEMHMPDKAFVKELV
RYAAPGRITIIYTMCHRNLSAGEBALQPREONILDKICTFPLPAMCSTDVYNILDS
HSLDIDKADWSENVAPFMPAVIRLALTWKGLVSLRSGMSIKGALTMPLMEGTRK
GVIKRGIITCKPL"

BASE COUNT 352 a 269 c 303 g 372 t
ORIGIN

Alignment Scores:

Pred. No.: 2,27e-91 Length: 1296
Score: 1147.50 Matches: 229
Percent Similarity: 75.43% Conservat: 32
Best Local Similarity: 66.18% Mismatches: 65
Query Match: 62.70% Indels: 21
DB: 8 Gaps: 4

US-09-857-613a-28 (1-350) x AY087138 (1-1296)

QY 8 ProthyllestercysillethisthPheargserinseProatqthPheala 27
Db 120 CCACTCTTCTTGGCTCAAACTCATCGCTTCTT -GGTCTCATCTCTCTCC 178
QY 28 ArgilearVal---GlyProarqsertrpIaProilearqalaSeraIaSer 46
Db 179 TCAGTCTCATGACGACCAACGCTGACGCTGCTGCTGCTCATACACT 238
QY 47 GluArgGlyGluIleValleuGluGlnLysProlLysLysaspLysLysLeuGln 66
Db 239 GAG-----GCTCTAAGA 250
QY 67 LysGlyIlelealapheryrAspGluSerSerGlyLeuTrpGluasniIeTrpGlyAsp 86
Db 251 AAGAGATACCGAGTCTCAATGAACCTCGGTTGTGGAGAGATTGGGAGAT 310
QY 87 HismethisthisclypethyrAspSerAspSerThralSerleuSerAsp-----His 104
Db 311 CATATGATCATGCGCTTTATGACCCGATCTCTCTTCACTTCTGATTCGTCAC 370
QY 105 ArgAlaIa 123
Db 371 AAGGAAGTCAGATCCGATGATGAGAGCTCTCCGTTCCGCGCTTACTGATGATA 430
QY 124 GluArgSerLysTrpProLysSerLysLeuAlaSerGlyGlyGlyGlyGlyGlyGly 143
Db 431 GAG 490
QY 144 ArgTrpLeuAlaLysLysPheGlyAlaThrSerValGlyIleTrpLeuSerProValGln 163
Db 491 AGAATCTTCTCTTAATTTGAGCGTGAATGCGATTCGATTCACGCCCTGTTAG 550
QY 164 AlaIa 183
Db 551 GCCAAG 610
QY 184 ValAlaIa 203

Db 611 GTTGGCAGATCCGTTGGATAGACCATTCGAGAGAGAGAGAGAGAGAGAGAGAGAG 670
QY 204 GluSerGlyLysIleHisMetProAspLysAlaLysPheValGlyLeuAlaValAla 223
Db 671 GAGAGGTGTAGATATGCTGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 730
QY 224 AlaProGlyAlaIleIleIleIleValThrTrpCysHisArgAspLeuGlyProAspGlu 243
Db 731 GCTCCAGAGAGATGAGATATATAGTACATGCTGCTCATGAAATCATGCTGCGGGAG 790
QY 244 GlnSerLeuHisProTrpGluGlnAspLeuLysLysIleCysAspAlaValTrpLeu 263
Db 791 GAAGCTTTGACCGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 850
QY 264 ProAlaTrpCysSerThrSerAspTrpValLysLeuLeuLysLeuSerLeuGlnAsp 283
Db 851 CCGCTTGGTGGCTCCACGATGATGATGATGATGATGATGATGATGATGATGATGAT 910
QY 284 IleLysSerGluAspTrpSerArgPheValAlaProPheTrpProAlaValIleArgSer 303
Db 911 ATTAGAGTCCGATTTGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 970
QY 304 AlaPheTrpTrpLysGlyLeuSerSerLeuLeuSerGlyGlnLysTrpIleLysGly 323
Db 971 GCATTACATGAG 1030
QY 324 AlaLeuAlaMetProLeuMetIleGluGlyTrpLysLysAspLeuIleLysPheAlaIle 343
Db 1031 GCATTGACAAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1090
QY 344 IleThrCysArgLysPro 349
Db 1091 ATCACTGCCAG 1108

RESULT 7

AC006193/c

LOCUS 118335 bp DNA linear PLN 28-MAY-2000
DEFINITION Arabidopsis thaliana chromosome I BAC F13011 genomic sequence,
complete sequence.

AC006193
VERSION AC006193.3 GI:4733953
KEYWORDS HTG.

SOURCE
ORGANISM Arabidopsis thaliana.

Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
AUTHORS 1 (bases 1 to 118335)
Federici, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
Altati, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P.,
Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
Luros, S., Schwartz, J., Shinn, P., Tortum, M., Vysotskaia, V.,
Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.

JOURNAL
AUTHORS 2 (bases 1 to 118335)
Federici, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
Altati, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P.,
Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
Luros, S., Schwartz, J., Shinn, P., Tortum, M., Vysotskaia, V.,
Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.

TITLE
JOURNAL Direct Submission

Submitted (09-DEC-1998) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA

REFERENCE
AUTHORS 3 (bases 1 to 118335)
Federici, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
Altati, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P.,
Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
Luros, S., Schwartz, J., Shinn, P., Tortum, M., Vysotskaia, V.,
Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.

TITLE
JOURNAL Direct Submission

Submitted (04-MAY-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA

REFERENCE
AUTHORS 3 (bases 1 to 118335)
Federici, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
Altati, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P.,
Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
Luros, S., Schwartz, J., Shinn, P., Tortum, M., Vysotskaia, V.,
Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.

TITLE
JOURNAL Direct Submission

Submitted (04-MAY-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA

REFERENCE
AUTHORS 3 (bases 1 to 118335)
Federici, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
Altati, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P.,
Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
Luros, S., Schwartz, J., Shinn, P., Tortum, M., Vysotskaia, V.,
Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.

TITLE
JOURNAL Direct Submission

Submitted (04-MAY-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA

Stanford University, 855 California Avenue, Palo Alto, CA 94304.

USA

REFERENCE
AUTHORS
4 (bases 1 to 118335)
Federispiet,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
Gonzalez,A., Kremetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
Linos,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.

TITLE
JOURNAL
Direct Submission
Submitted (11-JUN-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA

REFERENCE
AUTHORS
5 (bases 1 to 118335)
Federispiet,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Ecker,J., Theologis,A.
and Davis,R.W.

TITLE
JOURNAL
Direct Submission
Submitted (28-MAY-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA

COMMENT
FEATURES
source
On May 4, 1999 this sequence version replaced g1:4678192.

Location/Qualifiers
1. 118335
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/map="g1147"
/clone="F13011"
1092. 2311
/gene="F13011.1"
join(1092..1225,1474..2311)
/gene="F13011.1"
/note="Hypothetical Protein"
/codon_start=1
/protein_id="AAD38245.1"
/db_xref="GI:5042406"
/translation="MGQLIDTPKCTNTONEKRFISTEKLAFQRTAQTIVMKLQ
RRPTKMKDMKQSSPTMTMTIPIPDHSTGVREDAMQVCKPWLKENPYCCSSQ
OOHQHOKRAFIWDLNFKVFMSELGKADDEIKELADYERKARRAELMTIKLAK
DVEERARAEAEONKRLFKELSEKSEPMVKKRDLSEEROMHRLAEVLAEREVOMK
LMDARFLSEKSELSEANROGERENRMKPKIILERCSSPARRCRCEPQIKRGINP
PRVVRARISKEKSGKLECCOKVOLKILRQKTPRCPLPLSSPP"

gene
CDS
complement(3352..3963)
/gene="F13011.2"
complement(3352..3963)
/gene="F13011.2"
/note="Hypothetical Protein"
/codon_start=1
/protein_id="AAD38246.1"
/db_xref="GI:5042407"
/translation="MGNCILFGGLGDEEDLLIKIKSDGVLEFYSPTAGFVSHGFS
GHAFSAVDLLMKPLADHLVPGOSYLFPPVIVSDELKTFVSGCHVRNSSLAIT
PYRSLIDVNRHLKRSYTDVFSRNSHIFRROKERTRRRTSSKGMKIMVNLINTEE
LLQILSEDRNTNELIESVAAKGFISSITSSSENFSLVQT"
8107. 9734
/gene="F13011.3"
join(8107..8140,8222..8358,8440..8489,8562..8893,
8959..9062,9144..9219,9304..9734)
/gene="F13011.3"
/note="very similar to alcohol dehydrogenase: Similar to
g1:1143388 class III ADH, glutathione-dependent
formaldehyde dehydrogenase. [Arabidopsis thaliana]:
Location of est 94M1577 (g1:12757072) and 94M15XP (3')
(g1:3450565)
/codon_start=1
/protein_id="AAD38247.1"
/db_xref="GI:5042408"
/translation="METQCKVITCKAAVAMGAEPLVMEDVYKVPORLEVRILIFT
SICHDLISAMKGENEARAYPRILGAEAGIVESVEGVEGAMAGDHVPIFTGEGCD
CHVCKDGLANKEFRFVDMKVMVTDGTRFPTSKDNKPIYHFLNTSFSEYVDS
ACVIRKPLPLEKISLSCGYSKDCMNEGAMVADIDOPASTVAIFGIGAVGL
AVAEGRARAGASKIIGIDINPDKFQIGRAGISEFTRPKESKAVHERMEITEGVE

YSEECAGSIEALREAFELSTNSGVTVMLGVASPOLLPHEMELFOGRSITASVFGC
FKPKPOLPEPTIOGLOGLINLDLFTISHQLPFDINDAMQLHQKARCLLHL"

gene
CDS
complement(10308..11958)
/gene="F13011.4"
complement(join(10308..10729,10802..10863,11042..11125,
11208..11384,11461..11617,11723..11958))
/gene="F13011.4"
/note="Location of ests 86H377 (g1:2414136), 62B10T7
(g1:931264), 228J15T7A (g1:1217083), 89% identical to
membrane related protein Cps [Arabidopsis thaliana]"
/codon_start=1
/product="putative membrane related protein"
/protein_id="AAD38248.1"
/db_xref="GI:5042409"
/translation="MNGLSLVPTFESHGFDLIYFVAPLHINAVAGVGVWMPK
WAVLDSKLSNPKFLNLQLPSTLSKTHHNPVLSSEKTEGTVTDQFRLMKLVE
VNDGSPCWIMMDRSTPTFSYOAMRDPENGPPOVRSRTEVDPATREMDVDFWDE
RSKMDMLYSTLEKCDKDTGMVAVWVKRFPFCDSREYIIGRRIMDAGRFEYCITK
GVQYSPRONKPRRYDLVYSCWICIAVSKRQDGMTECEVILFHEHMGJPEIAK
LGVROGMGAIVKRIEGLRAYORAKAGAGLSPSAIIAHINIKRVSAREPMNERGSIAE
VTRDKPTGKNIKILVYGCAIALACTLDKGLTKAVIFGVARRFARMGRM"
13535..15717
/gene="F13011.5"
join(13535..13647,13808..14022,14535..14734,14807..15315,
15402..15717)
/gene="F13011.5"
/note="Identical to 'Tubulin Alpha-1 Chain [Arabidopsis
thaliana] (g1:135391) Location of ests 139J20T7
(g1:2762968), 214C14T7 (g1:1159108)."
/codon_start=1
/product="alpha1 tubulin"
/protein_id="AAD38249.1"
/db_xref="GI:5042410"
/translation="MREIISIHIGAGIOVNSCMEVLCLEHGIOPDGTMPSDTVGA
CHDAFNTFESESSGQHPVAVFLDLEPVTIDVVRGTROLPHRPOLLISGKDAANN
FARGATVREIYDCLERKLADNCTGIGLGVFVNAVGGSGSLLESLERSLD
FKKSKLFTFIYDPSQVSTAVVAVPVSVSTHSLBHTDVVLLDNEATYICRSLD
IERPTYSNRLISORTISLTSLSLRDGAANNVDTEFQNNVLYPPIHMLSYAVI
SSKATIHEDSPVEITTSVFEPSSNMAKCDPRKGTMACCLMIRGVIVRQVNTAVAA
IAKARTIQEVDPEPTFGKGINQPSVAVPGGLAVORAVCNINNTAVAEVFSID
HFDLTVYSRAFAHMYVGRGMEGSEAREDLAALEKDYEEVGGAEDDDEGDEY

gene
CDS
complement(16208..17133)
/gene="F13011.6"
complement(join(16208..16338,17040..17133))
/gene="F13011.6"
/note="Unknown protein: Location of ests 247E21T7
(g1:2413341)"
/codon_start=1
/protein_id="AAD38250.1"
/db_xref="GI:5042411"
/translation="MAEPPAAVAEYVKMDLFEDDDEFEFEETINEDWLEKEEYKVSQ
QWEDDDDDVDNDFSRQLKELENGDKK"
complement(17641..19522)
/gene="F13011.7"
complement(join(17641..18251,18334..18833,19281..19522))
/gene="F13011.7"
/note="31% identical to glucan endo-1,3-beta-glucosidase
precursor (1-3)-beta glucan endohydrolase) [Triticum
aestivum] (g1:1706551)"
/codon_start=1
/product="Similar to glucan endo-1,3-beta-glucosidase
precursor"
/protein_id="AAD38251.1"
/db_xref="GI:5042412"
/translation="MATHKLKPRTVVQMLKDNINIKVLFDADETTGALAGSGLVEM
VALINDOLKVTMSYDRAKMDVRRKNNVTRVNDGYNITFVAGNPEPLKSYNGSFINT
FPLANIDNANAGAGNSYKATVPRPLADVYDPSASVPSAGRFRPDIDIGMTQIYD
FLKNNAPITINIPFLSLTGNDPFLNTVAFPGAEITNDNGIDITNVDPANDTIVS
SLKAVGDMPIIVGEGWPTEDDKHANASAVRFYNGLLPRLGTKNPLRLPTYLEV
SLGLGDEDDKSIAPGFEERHMDIKFKEDGPKPPIIDSGGOSGFLGADNPVYLRNK
WCENFPAKDLTKLIANDIYACFESDCTALGYSCTTADANASAYAFMFEQVQNK
DESACYFQGIATITITQNIISGQCNFPIQIYVASSASSSCSYSLVYLIVWFLLSGMF

Oy 122 ----- 122
 Db 16914 TAGTACTGCAAGAAATACGAAAAAGACAAATTTGTCCTGATTCACGCCCTAGTTAC 16855
 Oy 122 ----- 122
 Db 16854 CTTGGAGTACGACATGTTCAATTAAATTTGCTAGCTAGTAAATTTGCTTGATGCC 16795
 Oy 123 ----GlnGluArgSerIleThrProLysSerIleValAlaPheValGlySerGlyIleGlyGly 141
 : : : : :
 Db 16794 GCAGATGATCCGAGAAACCCAAAGTGTGATGTTGGCTGCTGGCATTGGTGGT 16735
 Oy 142 SerSerArgTyrIleuAlaLysLysPheGlyAlaThrSerValGlyIlePheLysSerPro 161
 : : : : :
 Db 16734 AGCTCAAGATCTTGGCGAACAATACGAGCCCAATGCTACGCGCATCAGCTTGAGTCCG 16675
 Oy 162 ValGlnAlaGlnArgAlaAsnAlaLeuAlaAlaGlnIleLysLeuAlaAspLys----- 179
 : : : : :
 Db 16674 GTGACGCTGCAAGAGAAATGCGCTCGCGAGACAAAGGTTATACACAGAGTGGCGT 16615
 Oy 179 ----- 179
 Db 16614 ATTACTAGCTGTTATTCCTGTTCAAAAAAATTTCTACTGTTATTCGTATCGGAGATTAA 16555
 Oy 179 ----- 179
 Db 16554 TCTCGCTAGTACATGATGATGATTTGACATATACAGGGTTTAACTGCTGATAC 16495
 Oy 180 ----- 180
 : : : : :
 Db 16494 TCACTGCTGCTGAGACACAAATGTTTAAAGAGCTCTCTTCAAGTGGTGTGATTCAT 16435
 Oy 188 uGlnGlnProPheSerAspLysGlnPheAspLeuValIleThrSerMetCysSerGlyIleGly 208
 : : : : :
 Db 16434 GAGGACACCTTTTCTGATGCGAGATTGATCTGCTCCATGAGAGAGTGGCGAGCA 16375
 Oy 208 SmetProAspLys----- 212
 : : : : :
 Db 16374 CATGCCAGACAAACGCGAGGTAACTCTCTTTTTCCTTACAGAAAAAGAAA 16315
 Oy 212 ----- 212
 Db 16314 ATGGAAATGTAATCCCTCCGTTTACGTTATTAAGACTTTTACAGATTCGACGTTTAT 16255
 Oy 212 ----- 212
 Db 16254 ATAGATGTTAATGAAATAGTACACACATATATGTTTGAATTCATATATATATGAA 16195
 Oy 212 ----- 212
 Db 16194 ATGGGTAACTAGTAAAGTTTATTAATATGAAACGGAAGAAATATCTTGATCGGAGAAC 16135
 Oy 212 ----- 212
 Db 16134 GTTTCAAATCTAGCCATCAGGTTACGGTGGCAGTGGTACCATCGGTTGGAATCTT 16075
 Oy 212 ----- 212
 Db 16074 GGTTCAGCAATCCGATCTGTCGATTTAAATTTGCCGGTTGCGAAATGTAGAAAAACA 16015
 Oy 212 ----- 212
 Db 16014 GAGGATTAATGAAACAAATTCCTATTTAATAAATCAGATGTCATCAATCATCTT 15955
 Oy 212 ----- 212
 Db 15954 GAGAGAGACTGTGATTTGATGATCTACTGAATAGTTTGGACTTTTTCACCTCTTTT 15895
 Oy 213 ----- 213
 : : : : :
 Db 15894 AGCTAAGAGTTCTTAACCTGACTCTTTTACAGTTTGAAGCGAGCTGGCAGCGCTCGCAGC 15835
 Oy 224 aProGlyAlaIleIleIleIleValIleThrIlePheHisArgAspLeuGlyProAspGluG 244

Db 15834 TCCTGGGCGAGAAATATATCATTTGTGACCTGTCCATAGAGAACCTGACCATCCAGACA 15775
 Oy 244 nSerLeuHisProIleGluGlnAspLeuLysLysIleCysAspAlaTyrIleThr 264
 : : : : :
 Db 15774 GTCCCTGAACCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 15715
 Oy 264 oAlaIlePheSerThrSerAspTyrValLysLeuGlnSerLeuSerLeu----- 281
 : : : : :
 Db 15714 AGACTGCT 15655
 Oy 281 ----- 281
 Db 15654 AAAACTTTTCATGCTCTGACACTGTAAGTAAATTAATTAACAATGATATGTTGCA 15595
 Oy 281 ----- 281
 Db 15594 CATCAACTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 15535
 Oy 282 -----GlnAspIleLysSerGluAspTyrSerArgPheValAlaProPhe 297
 : : : : :
 Db 15534 ATGTGAATGTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 15475
 Oy 297 pProAlaValIleArgSerAlaPheThrIlePheLysGlyLeuSerSerLeuSerSerG 317
 : : : : :
 Db 15474 GCTCGGCTATTAATATACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 15415
 Oy 317 yGln----- 318
 Db 15414 TATGATCTGCAATCT 15355
 Oy 319 -----LysThrIleLysGly 324
 : : : : :
 Db 15354 AACTGAATTAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 15295
 Oy 324 lAlaLeuAlaMetProLeuMetIleGluGlyIleLysLysAspLeuIleLysPheAlaIle 344
 : : : : :
 Db 15294 CAATGCTGATGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 15235
 Oy 344 lPheTrCysArgLysProGlu 350
 : : : : :
 Db 15234 TCACCTGCTCCCAAGCCCGAA 15215
 RESULT 9
 AP003587 343550 bp DNA linear BCT 28-NOV-2001
 LOCUS
 DEFINITION Nostoc sp. PCC 7120 DNA, complete genome, section 7/19.
 ACCESSION AP003587 Ba000019
 VERSION AP003587.1 GI:17130808
 KEYWORDS
 SOURCE Nostoc sp. PCC 7120 DNA.
 ORGANISM Nostoc sp. PCC 7120
 Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 REFERENCE
 1 Kaneko, T., Nakamura, Y., Molk, C. P., Kurita, T., Sasamoto, S.,
 Watanabe, A., Iriyuchi, M., Ishikawa, A., Kawashima, K., Kimura, T.,
 Kishida, Y., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A.,
 Nakazaki, N., Shimo, S., Sugimoto, M., Takazawa, M., Yamada, M.,
 Yasuda, M. and Tabata, S.
 Complete genomic sequence of the filamentous nitrogen-fixing
 cyanobacterium *Anabaena* sp. strain PCC 7120
 DNA Res. 8 (5), 205-213 (2001)
 JOURNAL MEDLINE
 REFERENCE 2 (bases 1 to 343550)
 AUTHORS Kaneko, T.
 TITLE Direct Submission
 JOURNAL Submitted (02-MAY-2001) Takazawa Kaneko, Kazuo DNA Research
 Institute, The First Laboratory for Plant Gene Research, Yana
 1532-3, Kisarazu, Chiba 292-0812, Japan
 (E-mail: kaneko@kazuo.or.jp,
 URL: http://www.kazuo.or.jp/cyanobase/
 Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)
 FEATURES
 Location/Qualifiers

source 1. .343550
/organism="Nostoc sp. PCC 7120"
/db_xref="taxon:103690"
/note="synonym:Anabaena sp. PCC7120"
complement(94. .411)
/gene="a111719"
complement(94. .411)
/gene="a111719"
/note="ORF_ID:a111719
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB3418.1"
/db_xref="GI:17130809"
/translation="MRKKSATLEAVHETAEDLHKAGIMQTLREFEHLCLPPIEPL
EPQAIKEIRRESSVQAVFARILINISTVQKWEIGOKRPSGASLKLHLVKNNGLS
VLY"
complement(504. .782)
/gene="a111720"
complement(504. .782)
/gene="a111720"
/note="ORF_ID:a111720
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB3419.1"
/db_xref="GI:17130810"
/translation="MKIKNRTFDRMARKEGKLNLSICNAVNEMAPGLYDADGGGLF
KRRAKPGKSGSGFRLIATNNEVDGFLFLAQKTAATVLTTHKKKL"
935. .1507
/gene="a1r1721"
935. .1507
/gene="a1r1721"
/note="ORF_ID:a1r1721"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="BAB3420.1"
/db_xref="GI:17130811"
/translation="MRHKDNKKNQALICDAAIELITANGPADTSMSKIAKANYSPATI
YVFEENKEKLINKLYLEFQKQISAEILGVNONLSVEKAFKMINNNYKAIKPPVFE
APTEOFANSPMVRICKDESLYSKPEMLDLPNGKKEKVKDIPLEIFTAFTAPPLTG
LKEHFSGAILDEKMLETTYKIAMDAVTN"
1569. .2504
/gene="a1r1722"
1569. .2504
/gene="a1r1722"
/note="ORF_ID:a1r1722"
probable oxidoreductase"
/codon_start=1
/transl_table=11
/protein_id="BAB3421.1"
/db_xref="GI:17130812"
/translation="MKNEKNNAENIPQKRLAIYVSSSGICGTARVLANKQASVI
IAVNLDKGNKALAKIIOQKADADVKLMEJDLNLAIVSKNFAENFRNNYLRDLILIN
AGVMIPIYSKTTQFELQFGTNLGHFALGOLLEFIISTEGSRIVAVSSGANNMGCI
DFEDLMEQSRVAKKAYGDSKLANLFTYELDRKLDKDNGLDILVTASHGWTATELO
RTAGGIYKYLNGIVADODITMGALPTLAAIEAGLKGAEVFGPNGCFMNRGYPKIVESN
ELSKDQALAKKLVVSEKLLTDVKEEFENKKAQNSAG"
complement(2701. .3462)
/gene="a111723"
complement(2701. .3462)
/gene="a111723"
/note="ORF_ID:a111723
probable aldolase"
/codon_start=1
/transl_table=11
/protein_id="BAB3422.1"
/db_xref="GI:17130813"
/translation="MSVFSRPLPVPFNSIEEERLHRKORLAAAFRLGKFGFSGING
HITARDPEPTDHFNVGKRYFGHIRYSIDLVLNRDGVYKGOALLINQAPAIHSOIH
EARDVYAAAHASISYKAWSSIGRLIDPTODACAFYEDHALIDDTGYVLETSEBK
RLAETLGNKALILRNHGIITVGHIVDEAAFWYISLERSCOQLLAFAAQPKIIGHE

gene TARLTHSQGCEPKAGMFSFQPLYDRVAREEPELFD"
complement(3764. .5167)
/gene="a111724"
complement(3764. .5167)
/gene="a111724"
/note="ORF_ID:a111724
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB3423.1"
/db_xref="GI:17130814"
/translation="MSSTSGVPGCLDTHPTTELIOGSLITASLLVFLVIALGIA
IALMSFALRHNSVQQTIFIGEMAIRAQILRGLOHALVILLVGGFFSSTLSNRH
HMOQAKIAQVQSVSGDKLEQIAPOVRYVQEBEYTYNTQVNGKIVKYNDOQVSRLLS
LACSOIQVLEOSTDVQGRSAVRADYADYKVYNRLKDINNEFEIEQPNQYSLNS
KYVERORTLEPVPNGVSPAFLEGGQETSFVTVQAOGPRVYNASAGOLLSNRL
PTLIANPGADAFAGIVPKERKADGRSTQVYIMDDNVSVKNPRGCVFTNRPITQGTII
PRILLAPALFLMWILLIYLSLPSISLTNVAIASIYFACILITLITYSREDAQLAMTA
ISLVLLTLTWGLGSTRSASLAIITCTAGAVVPLGLVLPYSILTLVAGLSTVIMLA
VRHMYMNINWQPEDSKS"
complement(5492. .5902)
/gene="a111725"
complement(5492. .5902)
/gene="a111725"
/note="ORF_ID:a111725
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB3424.1"
/db_xref="GI:17130815"
/translation="MFLKPPSGLMVITASLTNGIAITYSGRSIAEAKSNFFLYQVGA
DQGFPIALDASINGTFTLVQOHGEGIAKRLHAACSGGRLETKFSLIYANGKLT
SEDTQDEIPEPKPTPDASMEIVCRVANNHQK"
6321. .6842
/gene="a1r1726"
6321. .6842
/gene="a1r1726"
/note="ORF_ID:a1r1726"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAB3425.1"
/db_xref="GI:17130816"
/translation="MKSQHLQATTLAGEVSQOYLKETALNAIAELQDFIDMCPDARE
VRKALAVKLYVQGYLDYDEIQKIDVSRGSITGKQAYEEGIDGLRKYGRKSYNA
KQOEVLISMLQTKDCWELGELBYEQALFEYDVVESQSYDLSSEAGISMKKTKVNP
KADENVAEKKRD"
6919. .7431
/gene="a1r1727"
6919. .7431
/gene="a1r1727"
/note="ORF_ID:a1r1727"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAB3426.1"
/db_xref="GI:17130817"
/translation="MWGDSGYGWSKDOEIAVPVNERDKORYYGAVDYLIGEIVLK
AYDAGSKNTINLLEVLANSPOQRLLIFPDGASTYRSKRTIRFLDSVNSLPTEDQK
IHCVRFPNCPVONPIEDIWLOAKTWRRFCALIPSSHLKWIIEWIRHTTDFDPL
QMYGTYSKIR"
8064. .9401
/gene="a1r1728"
8064. .9401
/gene="a1r1728"
/note="ORF_ID:a1r1728
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB3427.1"
/db_xref="GI:17130818"
/translation="MTGNARNAHAFRLVSGNCAFGSGSRYSLSKVEYVIGRDPSCQ
VLIDAMYRMVSRRAHVVRVAVSSVSKFSWVLCIDLSANGTYLNGQRLYGCQDELHAG


```
/gene="SB234M12.3"
/note="highly similar to Oryza sativa gene for putative
receptor kinase GB:AF128457 GI:4680335; similar to Oryza
sativa receptor kinase-like protein gene GB:AF119222
GI:4680488; similar to Arabidopsis thaliana receptor-like
protein kinase emb:AL138654 GI:689879"
/evidence=not_experimental
complement(join(<12094. .12725,12823. .>14263))
/gene="SB234M12.3"
/product="putative receptor kinase"
complement(join(12094. .12725,12823. .14263))
/gene="SB234M12.3"
/codon_start=1
/evidence=not_experimental
/product="putative receptor kinase"
/protein_id="AA094320.1"
/db_xref="GI:22208505"
/translation="MTLLAFRLPPELLLAATAAANDNGNTTEAFAAALVNL
KSTPTDPTGLEAASATSPPPCDASPPGVCQKSLVGVRLTHMNLSTDFGAI
AKLRLHSVNLKRNALISGLPASLGLRLALYLSNNEGPIPADVFAMRLKLI
YLDNNRITGLPADAIKAPRLMELHDBNOIDGPVPEKLPESLKRNVSHNLSGTI
PQSAERYDASSFAGNPGICSGSDAACAAPALPPAPPPADADYMADEETSV
FVYIGILVILVLSGAVMLRDBDRNSAVTANDYAGSTTTGGAGAGAGSSKVA
DAASTTAAPRAGAVADMASSGSSSSHGCGSSGRRKEEVLNEDITAFGLPDL
MKASAEVYGNCTLGSAYKAMRNGVYAVKRLRDMNRFGRDEQHVHMLQLRHPNV
LPVGYHYRKEEKLIVSEYMRGSLYLLHGDQSPNRYITDMQGRVIRAVGVGLAF
LHERLGIAGRLVSMGDGADAPPPEPHGNLKGSNITLGDMEPLVDYGFPLVA
AQAPQAPAFSPGCTGTGGVARSADYVCLGVLELVTGFRPSQYLLSAGGDDVH
WAATAVAEGGAEIYDPAITAAAGDAVRLRVRGVCASPEPECRPSVAEAAWVEEI
GSAAGAS"
12558. .12578
/rpl_unit=c99
complement(13145. .13307)
/rpl_unit=c99
13490. .13517
/note="GC-rich"
14190. .14216
/rpl_unit=c99
complement(<15173. .>20442)
/gene="SB234M12.4"
/note="Similar to Arabidopsis thaliana gene for putative
sugar transporter protein GB:AY094465 GI:20453188"
/evidence=not_experimental
complement(join(<15173. .15605,16042. .16226,16382. .17055,
17854. .18423,18531. .18650,19447. .19501,19596. .19649,
20293. .>20442))
/gene="SB234M12.4"
/product="putative sugar transporter"
complement(join(15173. .15605,16042. .16226,16382. .17055,
17854. .18423,18531. .18650,19447. .19501,19596. .19649,
20293. .20442))
/gene="SB234M12.4"
/note="AT4935300/F23E12_140"
/codon_start=1
/evidence=not_experimental
/product="putative sugar transporter"
/protein_id="AA094321.1"
/db_xref="GI:22208506"
/translation="MSMTHCKLNVAQEDLMGATGAGDVQMCNGKLEEVAVLCAH
RIMAEVSTHOEVCMMYFAGAAIASRLIDCEAYLRTEFGFIVCKRAVLDEM
STVFSAVVYISGYAMGDFALLLENNHMKFELLNPSLEGITLAASTGALYIT
IFSGALMDLGRRAILVYSVLFGSGVLMMSPNYIYLRLRIVGSSGLVFCVP
YITSETSPNNRGLCTMPQMFICTITSYCLITMLTMSPPNRIMICAFAPSVY
YFALLVYIIPESRMLASDGKISEARVSLQRLGKHHDVSGEIAVLVDVDVITDSAV
GPARAOSFSGTSRTPWPRSTYMWLSDLVLDLGSIHNMSEGGSRNSFFPVEFSF
SEPEHNMNEHNDGSDQOTREYSAGVNNNGSLASLQSAEVEANDANTSPSE
GSSYLRHGTGSLAOFEMASIHDDIEEESEHGVLLHQSAPRVEESGRHOVRL
SETAPMKKRVILQCGIRALCYGMLTQALQOSTGICSLRARPILLPOVGYLSD
IGLSPHSTILSTDAALILPCTRAMLPMDCGRVYLATPPLILSLSYMSMSI
LENMGPDRALIVFHFALVFCSYVIGLDPILDLSEIPLTARAFSCFSLSWF
GGLLSAYCCLPVMMSTIGLGGAGCIVALVCCAPLFLFYRIPIETKMLLELLEFLS
ROBYVO"
complement(22125. .22161)
```

```
/note="AT-rich"
complement(22424. .22446)
/note="AT-rich"
repeat_region
23019. .23089
/note="AT-rich"
repeat_region
23130. .23172
/rpl_unit=ta
23495. .23526
/note="AT-rich"
repeat_region
26812. .26900
/note="AT-rich"
repeat_region
37014. .37038
/note="GC-rich"
complement(37399. .37462)
/rpl_unit=a
<37578. .>44288
/gene="SB234M12.6"
/note="Similar to Nicotiana glauca plumbaginifolia ubp1 gene
emb:AJ272011 GI:6996559"
/evidence=not_experimental
join(<37578. .37721,38745. .38806,38998. .39100,40241. .40317,
40518. .40579,40690. .40777,40913. .40988,41080. .41110,
41205. .41318,42137. .42222,42364. .42537,43066. .43258,
43335. .43537,43864. .43887,43980. .>44288)
/gene="SB234M12.6"
/product="putative oligouridyate binding protein"
join(37578. .37721,38745. .38806,38998. .39100,40241. .40317,
40518. .40579,40690. .40777,40913. .40988,41080. .41110,
41205. .41318,42137. .42222,42364. .42537,43066. .43258,
43335. .43537,43864. .43887,43980. .44288)
/gene="SB234M12.6"
/codon_start=1
/evidence=not_experimental
/product="putative oligouridyate binding protein"
/protein_id="AA094322.1"
/db_xref="GI:22208507"
/translation="MHPRMKQAAQAQQLMQQALLQQQQAQAPPLPFGHHHPGIL
AAPQEPVATSGMLPPGPDSTCRSVYVGNISLQVDTYLAQEVQSTIGVEGKLRIKE
KSSFGFIDYDRYALALILSLNGRPLGQPIKVMATYSTREDTSGHNFIVDLC
PEVDTATLFAFSGYSTCSDAVARMDQKGRSGGFVSFMQQAQASAINDLQKWL
GNROIKNMATKGANAGDEKQIVDSVDLTNNTSSESGKPNEDGESNPOFTYVYG
NIPHEATNNDVHLFFHSLCAGSIEEYRTRDQFGFVRSTHEAAQAIOMANGOLIG
GQIKITTYINGYKLRAYESDPCEVLYKWSRLVTLVAVYASVFLLENDANRO
GIDLILYASTAGFAVCGVLGVLGYLYLVLAQSDPDELRLMKGAQACRACVIRK
SLICGVSVECAKEGEAVAVARCTENELGMSQSKPTGTSASPPLPAPALFSTVG
SPTDLAYORMLNATMPNSLMGIALNQVALMGGRASQSIYDGFQGISMAATY
SNOQOQOQOQOQIMY"
37580. .37668
/rpl_unit=c99
complement(39121. .39142)
/note="AT-rich"
44247. .44274
/rpl_unit=c99
complement(45252. .45289)
/rpl_unit=c99
<45290. .>46757
/gene="SB234M12.7"
Alignment Scores:
Pred. No.: 4.17e-39 Length: 108553
Score: 579.00 Matches: 160
Percent Similarity: 38.90% Conservative: 31
Best Local Similarity: 32.59% Mismatches: 61
Query Match: 31.64% Indels: 242
DB: 8 Gaps: 5
US-09-857-613a-28 (1-350) x AF527809 (1-108553)
QY 94 AspSerAspSerThrValSerLeuSerAspHisArgAlaInlIleArgMetIleGln 113
DB 94288 AATTCACCTGATTATATAATTGGTTGAC-----ATCGTTCGCAACTTTGA 94241
QY 114 GluSerLeuArgPheAlaSerValSerClnuArgSerLysTrpProLysSerIleVal 133
```

Db	94240	CTACGATTTCAATGCTTATTTCTCTGTGCAGATATATCTAGAGAAAGAGGCCCAAAAACATATGTC	94181
QY	134	AspValGlyCysGlyIleGlyGlySerSerArgTyrGluLeuAlaLysLysPheGlyAlaThr	153
Db	94180	GAGCGAGGATGTCGCAATGAGGAGGAGCGAGCTCAAGGATCTGGCCAGAAATATACGAGCACAA	94121
QY	154	SerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaAsnAlaLeuAlaAla	173
Db	94120	TGTACGGGCATCAGCTTGAGCCCTGTTCACCTGAGAGAGAAATGCTTACTGCCGCC	94061
QY	174	GlnGlyLeuAlaAspLys	179
Db	94060	CAGGCTTTGTACAGACA-GGTGCCCAAACTCCGCTCCTTAATTAATCGTGATTAATG	94002
QY	179		179
Db	94001	TACTATTATTATCATCATTCATAGTAAGTGGGCTAAGGAGAGACTTGTTCATATCA	93942
QY	179		179
Db	93941	GTCGCTCTCCACACACACCTCTTTTCAGTTTTCACGAGATCAGTAACAGATCGCTATG	93882
QY	180	---ValSerPheGlnValAlaAspAlaLeuGlnGlnProPheSerAspGlyGlnPheasp	198
Db	93881	CAGGTTTCCCTCCCAAGTTGCTGATGCTCTAGAACACCTTTTCTCGAATGGAATGTTTAT	93822
QY	199	LeuValIrpSerMetGluSerGlyGlnHisMetProAspLysAlaLys	214
Db	93821	CTTGATGTGTCATGAGAGTGTGAGACACATGCCGGAATAAAG-AAAGGTAACTGTTT	93763
QY	214		214
Db	93762	TTTCTTGTGTTTTCCTGAACACTCTTAACCTTCAGAAATGACCAAAATTTCTTA	93703
QY	214		214
Db	93702	CTCGTTCGACGTAACCTGCATATGTTTGAACCTTACCTTGATGTGATTAAGTTTAA	93643
QY	215		215
Db	93642	TGCCGTAATGTCATATGTCGTAACCTCTTAACCTTACTGTCGTATACAGTTGTTAGT	93583
QY	218	GluLeuAlaArgValAlaAlaProGlyAlaIleIleIleValThrTyrCysHisArg	237
Db	93582	GAGCTAGCGCGCTGCGAGCTCTCGAGGAGCAATATCATCTGACATGCTGCCATAGC	93523
QY	238	AspLeuGlyProAspGluGlnSerLeuHisProTyrPglGlnAspLeuLeuLysLysIle	257
Db	93522	AACCTCAATCCCTCGAAACTTCCCTGAAACCCAGTAACAGACACTCTTAAGAGGATA	93463
QY	258	CysAspAlaTyrTyrLeuProAlaTyrPysSerThrSerAspTyrValLysLeuGln	277
Db	93462	TCTGATGCATACCTGCTTCCGACGCTGTTCCCTACCTTCAGACTATGTCAGCATTCGTAAG	93403
QY	278	SerLeuSerLeu	281
Db	93402	TCACTGTCTCT- CGAGGTAAACAAAAAAGAAAGAGTTCGTGATGTGATCTGAA	93344
QY	281		281
Db	93343	TAAATTCGAATGAATACGAAGACCTTGCTATGCTTACCTTCATTTGCAATACGCTTT	93284
QY	281		281
Db	93283	GATCAATATCTTGATGATTAAGATTTGAATGCTTCATCTACCTAGGGGACACCGCAAT	93224
QY	281		281
Db	93223	TTTGAAATTTGTTGCAATGATGTTATTTTCATAGAAAAAATCTATTTTGTGTTCATG	93164
QY	282	-----GlnAspIleLysSerGlnAspTyrPseArgPheValAlaProThrTyrProAla	299

Db	93163	TTGTGTACGATATACAGACAGCTGACGTGCTGGAGAACGTGTCCCGTTTGGCCCGCT	93104
Qy	300	VallIleatGserAlaIphethrTriPlpYsgIlyleuSerIleuSerSecIly-----	317
Db	93103	GTGATAAACCTGCTCATTTCTCATGAGAGGGCCTGACCTCTGCCTTAACATGTGTATAT	93044
Qy	317	-----	317
Db	93043	CTGACACTTACTCGCTTATCTTCATATAGCTCTTTTGCAGAACATGGACCTGCACCAAC	92984
Qy	318	-----Gln-Ly	319
Db	92993	AAATCTTAGACGGCAGCAGGATTTTATCATTCATCATGTGGCCCTCTGTAAGATGGCA	92924
Qy	319	StrIleYsgIlyAlaIleuAlaMetProIleuMetIleCugIlyTYrLYsAspIleuI	339
Db	92923	GACGATCAATGGGGGGATGGGATACTACTGCTGATCCAGAGCCTACAGAAAGGGCTCAT	92864
Qy	339	eIysPheAlaIleIlethrCysArgLYsPro	349
Db	92863	CAAAATTCACAGCATCACTGCGGAAGCCT	92833

RESULT	12
AP005375/c	
LOCUS	298750 bp DNA linear BCT 17-AUG-2002
DEFINITION	Thermosynechococcus elongatus BP-1 DNA, complete genome, section 7/9.
ACCESSION	AP005375 BA000039
VERSION	AP005375.1 GI:22295445
KEYWORDS	
SOURCE	Thermosynechococcus elongatus BP-1 (strain:BP-1) DNA.
ORGANISM	Thermosynechococcus elongatus BP-1

REFERENCE
AUTHORS
1
Nakamura, Y., Kaneko, T., Sato, S., Ikeuchi, M., Katoh, H., Sasamoto, S.,

Watanabe, A., Iitiguchi, M., Kawasumi, K., Kimura, T., Kusuda, T., Kiyokawa, C., Kohara, M., Matsunoto, M., Matsuo, A., Nakazaki, N., Shimo, S., Sugimoto, M., Takeuchi, C., Yamada, M. and Tabata, S. Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* Bp-1. *JOURNAL OF MOLECULAR BIOLOGY* 356: 1-12. 2002. In press

FEATURES

AUTHORS Kaneko, T.

TITLE Direct Submission

JOURNAL Submitted (05-JUN-2002) Takakazu Kaneko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research: 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0812, Japan

(E-mail:kaneko@kazusa.or.jp,
URL: <http://www.kazusa.or.jp/cyano/Thermo/>,
Tel: 81-438-52-3935(ex.2338), Fax: 81-438-52-3934)

Location/Qualifiers

source

1. .298750

```

/organism="Inermosynechococcus elongatus BP-1"
/strain="BP-1"
/db_xref="taxon:197221"
/note="BP-1"
complement(61..1443)
/gene="henn1"
complement(61..1443)
CDS

```

```

/gene="nema1"
/note="ORF_ID: c111720"
/codon_start=1
/transl_table=11
/product="oxygen independent coprophorphyriogen III"

```

```

oxidase_1"
/protein_id="BAC09272.1"
/db_xref="gi:22295446"
/translation="MSVVOIPDFDALLQKYRPLPRYSYPTAAEFTTFEDRHCFQI
DLRSENAEHLPSLTVIHPEFCOTACYPCCGCAIIYILOSOTIATYIDALAAETAAVA
LDLRPRVYOMHMGSGTPNVLTIDOVESLRRTLTHOHELENDISEVNPRIINDYD
LIRLHNGENRISFEGLODDPOVALNRPQPTWLPOVWEHRAAFESVNDILIG
PYTQSEARTAKTLRDPDRRIAVFNAPYLPNLPELQRIIDPTLPPSARVTLTIQ
VIERLISGIRKIGMDHRAKPTDELAIAOENATLKRNPQGYTTLTALIGGLTISI

```

gene
complement(1482. .2213)
/gene="ho2"
complement(1482. .2213)
/gene="ho2"
/note="ORF_ID:tl11721"
/codon_start=1
/transl_table=11
/product="heme oxygenase 2"
/protein_id="BAC09273.1"
/db_xref="GI:22295447"
/translation="MTVSLALREGTAEAHTLAHTAPMKCFVGLITSRLRLOFLAN
LYFYVDTLEAASNGDAVRAIYFPELNSGOLAODLAYGGEWDQVTPLATATY
YLSRLHDVAOSEPLLIATHSYTRMGDLSCGASLKTYHSQMLPECKGATFYEPL
PTBEDKAFKORTATIDSGDEBETARTVQENAFALNCNLMALESEBLKTTIGD
QWQSVNIENQPSRQHPREAIAT"
2486. .3550
/gene="tl1722"
2486. .3550
/gene="tl1722"
/note="ORF_ID:tl1722"
hypothetical protein
/codon_start=1
/transl_table=11
/protein_id="BAC09274.1"
/db_xref="GI:22295448"
/translation="MVAIAPRHSGSVKTIITREMLTPREYTTDFEQAKLDSROE
OLEMKLMDRADNRHHFVRDSEKGTWDLIDQTRAFIDYERSCVSESGFLRK
ELSRKLMDRNPFLAEIPLMARDEARHAGLNMDFGAMDLGYLSKRTYTFPL
PMWLYVYVLESEKIGYMYRIIYRLHEKHPESEFYPLNYFERMCCODENRQDILKAI
HSOPOLIOGGAKEAMREFLTIVATHTTLVYHERAKFYEALGDATAFDREVIKATHE
TAARTFVNLVNDHPEFYSLQRCEVLSNKLQALEATRHQPKWLKALKPLHQALIAH
LRLRYLPPYNAQOENCTYH"
complement(3547. .3858)
/gene="tl11723"
complement(3547. .3858)
/gene="tl11723"
/note="ORF_ID:tl11723"
hypothetical protein
/codon_start=1
/transl_table=11
/protein_id="BAC09275.1"
/db_xref="GI:22295449"
/translation="MLSRTPLIERLAEIGRDIYMDYAKMHLYLGDAKIATPLAEFL
PLLEQGEMLAQYTSVLDQEVSVPLGGDQFLSLERLLPKANQELIVELNRFQEELE
S"
gene
complement(3963. .4382)
/gene="psad"
complement(3963. .4382)
/gene="psad"
/note="ORF_ID:tl11724"
/codon_start=1
/transl_table=11
/product="photosystem I subunit II"
/protein_id="BAC09276.1"
/db_xref="GI:22295450"
/translation="MTTLGQPLPYGSGTGLSAADTEEKYAITTSPKQVEMPT
AGAAVNREGENVLFARKEQCLALAAQQLRPRKINDKIRIFPDGETVLILHPQGV
PEKVNKREAVNSVRSIGONPNSQLKFGKKRPD"
4763. .5398
/gene="tl1725"
4763. .5398
/gene="tl1725"
/note="ORF_ID:tl1725"
hypothetical protein
/codon_start=1
/transl_table=11
/protein_id="BAC09277.1"
/db_xref="GI:22295451"
/translation="MYLYLHGFASSSPRSSKAQYFRDRFEELGQPLLPIDINQDFPH
LTLRSRIQVEYALLTPEEPITLIGSSFGGLTAAWLAKKHQVQVLFLLAPAFEPALIH

gene
complement(5383. .6369)
/gene="tl11726"
complement(5383. .6369)
/gene="tl11726"
/note="ORF_ID:tl11726"
/codon_start=1
/transl_table=11
/product="delta(24)-sterol C-methyltransferase"
/protein_id="BAC09278.1"
/db_xref="GI:22295452"
/translation="MSHLGLITVVAVALVGLATLYLPPRKYESRSVAESYQNT
KDGILFEYMGCEHILHCHYGLPPPKDFRQAKVDVFEHVMWAGIDRLPPTGVLDVGC
VWSITENGPRMPKQCFARKELLYLKRGGTLLVADNQRDROQPLNFMRLITROLL
DOMAHPRFASIEGFEALANTGLVAGVMTADTQETLPSMLISMOGITYRPEGLIR
GLPGLVSLREVPFELMLRIAFMGICRFQMFRAVRAETPAVSLPEAPQVNC"
6619. .8262
/gene="r1ba"
6619. .8262
/gene="r1ba"
/note="ORF_ID:tl1727"
/codon_start=1
/transl_table=11
/product="riboflavin biosynthesis protein R1ba includes
GTP cyclohydrolase II and 3,4-dihydroxy-2-butanone
4-phosphate (DHBP) synthase"
/protein_id="BAC09279.1"
/db_xref="GI:22295453"
/translation="MPVFDSESLALRGEVIVVDENREHEDLIGAERVTPA
MINMAVHAGLIDCLAMEGRDELPLMTVNTDSNQATFYSVDAGARWYTG
SAEBRATVAGLIDPSTQPODLRRGPHVPLRSRGGVLRKAGTEAVALDTRLAGY
PAGYICETONPDGMSRLPOLMEYARTHOLKITISIALISYRLQHEHFRVREAAKIL
TGEREPIQYCRNSLDSSEHAIYKGPATRGQPVYRVHSELTDAIGSLCDPR
MOLQALKMTNAGRGVYVLRQGRIGLVNKRATSLDDGLDITYEANDHLSFPD
LRNKGVAQILNDGVKQIRLITNPPKLAGLKGYLEVVDVPLLEAVPPYNYLT
TKAEKILGLITQTYLITVAFRMQSOGLAADRYERLEKRLHLAGVHLIVREEARPYA
QALRGHPDLIVFEGFDQPHVADRQWYQPEHPYVLAQVTLRQCCQPTLKGFELVA
TGTPMLNLPMHDDROAQYTOQSPSEMRPHLIYSNAAG"
complement(8267. .8455)
/gene="ts11728"
complement(8267. .8455)
/gene="ts11728"
/note="ORF_ID:ts11728"
unknown protein
/codon_start=1
/transl_table=11
/protein_id="BAC09280.1"
/db_xref="GI:22295454"
/translation="MDVRTDKQCHNSVIVYWKSLAPLNKRVTRHPEKGVNLGVVP
TPQSORCCOYSQTSIGAF"
complement(8456. .9043)
/gene="tl11729"
complement(8456. .9043)
/gene="tl11729"
/note="ORF_ID:tl11729"
similar to hct1 protein
/codon_start=1
/transl_table=11
/protein_id="BAC09281.1"
/db_xref="GI:22295455"
/translation="MMDCPLPIVVPQMBAPHPNLTDSSALHLMLTLPPDPVPGYI
LRATIRYQPNLGERPLPBAAGKPYNLGLFPMNSHSGNLAVLASGRAVGVDEL
RLCFQRAISRFRFGAALQDRILEGGRSLEQANTTYEAMLKAQDGGIGWORTAASLE
HWVASFPGDAIASVVVLTTPPOCFEFPREANA"
complement(9050. .9559)
/gene="tl11730"
complement(9050. .9559)
/gene="tl11730"
/note="ORF_ID:tl11730"
hypothetical protein
/codon_start=1

gene
CDS
/db_xref="GI:17131112"
/translation="MKTPTLDMNEAVEGCAALGLITLGYGRVPLPOLRRCGSRD
GSKASNMVAARTYGLAEKFKELKQVQQLPPYIVENFNHFWVEEGFQAQVYIN
DPATGPRHVSIVFEGYGLVLMVPGEGFQKGRKPSMIASLRGASGALVYC
IVAGEFLTIGLIVFESQIEVDLIVGRMLPLHLLAMAIAALISGLTLRLRY
LRLKILRLAVGSSREFMHLILPVSFYQRMFAGELSNRTTLDNOVADLAQLTTL
IDTLVLFALVMQYDWTLMVVSFAVAVLTLOMISRSRDIANOPLIOYGYAG
ISIALOSIETLKASGLSEDFSSRMSGYTKRLNSOEGLGTTNQLEGNLIRLDVA
MLLIVGGFRVMDHSLIGMLIAFGGLMSGFQGPVNNLVNFGTTQLEGNLIRLDVA
LDNPGVEGKTSSSSLQALPKLQGVLELKNITFGYSRLAPLISFNLSIKPGQVVA
LVGSGSGKSTIAKLVSGLYCPMTGEIYFDSQPREISQLLTNSVAMVEGLLEGG
TVRDLNLTMDTVPDPKMLRACODAAIADVIFMSGGYDAELLEGANLGGORLE
IARLVNNSIILVMEATSALDAETEKITIDOLRLRRGCCITVAHRLSTIRCCDIIV
LERKVVQAGTQOELMQVEGATSRILRTEG"
/gene="a112022"
/complement(6028..7395)
/gene="a112022"
/complement(6028..7395)
/note="ORF ID:a112022
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB73721.1"
/db_xref="GI:17131113"
/translation="MOVSPKSMPLVNLGSLVGVAFIMSTIGSTPIYEGRGVLI
SNVPLQSKSAGQIMALNVKVDVIAKKQVATTDQAEIRKQLQQRKLTQESODK
AVSLQGRLEDEERSQOQROYLRQLILELAIPLTKNTSVIQROQLQRLQ
QAQLTAVFLRRMKIROELFOKEGASIGSEALKAQOEYLNQLEKIDADQALKEIDV
ETQOEKDYRENTLVADLOSQKQLODQSAVQODLENTARNKEIOEVRKEISLE
LQLSNSOITISQHSRILEITLPGOVNAGTRATIEANFOSKLVGTVYPAVEAG
KIQPGMSIOTPPQVKKRERFEGIVGNVTSISRFPTKEAANEVNSEVLELVQOO
GLIOYFSLHLDANTPSCYKMSSTGPHLNISSGTTTYRVAVRERAPITFVPLTRS
ASGTY"
/complement(7472..9502)
/gene="a112023"
/complement(7472..9502)
/gene="a112023"
/note="ORF ID:a112023
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB73722.1"
/db_xref="GI:17131114"
/translation="MSDYCLVMPYAVERBSIALGLKARLQAKLEPTLIYPIQF
AEICLDVYAMISENFIETFLSEMTFSGVLEPEFNPNSYFDLISPCFKSEYLPD
RNLEKFAVAREQATAFIESVAQSVANLQPLRIGCSSTFOHCASIALKRYRELAPE
ITTFMGANGCEGAMGLATHTESVWDFTISGEGDLFNEFCYQLDGRVEDPSILPY
GVISPSHRTSTKISIAAPRASVQNLDOVPIPDYDIPOTIAHSTLSPYIOEGLIETS
RGCMWGGKHCHFGICGANGSIYRSKSDPVIAKERTQOCERGLKRFQVONITLDMN
MNTVLPVLAQDEANITYETKANLRWEHLQQLAAGVRCIDPGIESMHNSTLKLANK
GNTAMMNVQLKWARELGIIVFNITLADPPGSDMYEMEMFLPLIAHLQPPSGASE
IRYDRFSPIYERPSDMGNLSELRPKYVYPAPLISELAFEPDONNLVRYATEE
NEQTSKTERPLENLKLMQNMRLQSLYKHKPELPLITLSMDSGLTITIDSR
PCATESLFLBGLAVWYKVCDRALTYREVSIEIKKPELPLDVPMSDEIPIYSLEARK
ILLDIDGRLLSALGSCSPYISVSELPAGITIDLHEFRHNSKSEKLLRP"
/complement(9616..9909)
/gene="a112024"
/complement(9616..9909)
/gene="a112024"
/note="ORF ID:a112024
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB73723.1"
/db_xref="GI:17131115"
/translation="MSLVMTKELEEKIIVKAGODESFRKQALVNPSTLEKGINLP
SSIDVNLVETPGRNLSLOALPNSBODTISEALEVSAGRISGEVSYVYKVF"
/complement(10195..13482)
/gene="a112025"
/complement(10195..13482)
/gene="a112025"
/note="ORF ID:a112025
similar to Mrsd protein"

gene
CDS
/codon_start=1
/transl_table=11
/protein_id="BAB73724.1"
/db_xref="GI:17131116"
/translation="MNVSKOSILIKTIERSNSIRSLSSSEFIENDROINANONAKLEK
WYOFIAKNOHNEFEKRLMDGDTITGSCVGNLVNLABEKIYNNVETLKGMOYATF
VTPENSNKFLQDQHRHLPDRFPFEETLIVPVDAKKKLIESAGSSTYLLSONKHTQ
MERSLLEHDTGICISPLETESIFRSLSSPVLRINOLQSGCSQOYTDVNNKLVQ
GLSFFQESVILARLIATSVDFVWATGEFSLRASDSQKIQLTQFQPKLQOVVYQ
LAMSDFHNGRAVIVIKFASGKLVYKKEPSLLEKAYDEFLDMINOKKINLPKLKI
LISNVMGNEFAVAILPCODQVAVHYORAGVLCIYVLKGNDCHEMLACGGQPV
LVDLETLAHPRTWLSKDDADKASINELQDSVGTGCLPGQOIIIPYQETELKDFP
GLGGGQEDMPFRANKKWHINTDSVAVIYRETKLPLPKNRPGEINTDLSLNHSELI
DGFROMYFLOQKKEELASDSPTIAFESNOKRVLRIRNTYVASILQNSLNKCLTGT
VEHSIQDLDSRAFLSSBDBKPLMLAELKQALBQDLPYTFAYSQNSVNDIISSEOT
INKEFLSSSYDDVIAHLROMDADLAKOISLRGSLYCLTEQNSLPLEPNLMD
IYAPIERGMLOQAIAIQOOLEQOARIGKDNSVWIMGMYGEAEORFQLOPVGNSLYD
GCCGSLFLAALASVTNNAERGESLASLOSRTKLODPHIDLOOKLIGOMIGGATG
LASIYVALYRISSEFGEVBLIDQDQALASMTYNTATYNOAGVMTGAGTTLTLLAL
YKATEPATLQQAISMGNLTLDTRVTVDEVYQAMNAEKVLGFSHGAAGIAYALQ
LYAATDEPIFASAREATTLBQRLFSLSNCDLDVLSIDGSLSTTEKSGSLGIS
VARLCTNLIDPDEIROELIDNLHTIHSSIQVLDNICCGNFSHTEMLIVASQOLSH
ELLETVANNRATQVLSRANOIGYFQLPHLPADVVSFGCGQIAGYIGELLRLSCDDL
PSILIMO"
/complement(13877..14227)
/gene="a112026"
/complement(13877..14227)
/gene="a112026"
/note="ORF ID:a112026
similar to nitrite hydratase"
/codon_start=1
/transl_table=11
/protein_id="BAB73725.1"
/db_xref="GI:17131117"
/translation="MSBOTKFKREIEAOIYOAMKDETYROELLNNSKAVIEREPAIO
LPBEINHVAVEENDSNFYFVIPARNLEDELSEBQLEAVAGSGIDLFTFYEKTYDR
LKDRCKEIKAYVD"
/complement(14385..14720)
/gene="a112027"
/complement(14385..14720)
/gene="a112027"
/note="ORF ID:a112027
similar to nitrite hydratase"
/codon_start=1
/transl_table=11
/protein_id="BAB73726.1"
/db_xref="GI:17131118"
/translation="MSBOTKFKKDVENAOIYOAMKDEYROELLNNSKAVIEREPAIO
LPBEINHVAVEENDSNFYFVIPARNLEDELSEBQLEAVAGSGIDLFTFYEKTYDR
LKDRCKEIKAYVD"
/complement(14854..15042)
/gene="a112028"
/complement(14854..15042)
/gene="a112028"
/note="ORF ID:a112028
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB73727.1"
/db_xref="GI:17131119"

Alignment Scores:
Pred. No.: 3,856-22 Length: 334520
Score: 392.50 Matches: 102
Percent Similarity: 51.70% Conservative: 50
Best Local Similarity: 34.69% Mismatches: 101
Query Match: 21,458 Indels: 41
DB: 1 Gaps: 12

US-09-857-613a-28 (1-350) x AP003588 (1-334520)
QY 69 IIEAlagLuphetYrAspLgU-----SerserGlyLueTrpGluAsnLIeTrpGlyasp 86
DB 127499 GTAGCCATTCTTACGACGAGTCTGACTGACGAGCGATATTATTAGAGTTTACTGGGGCGAA 127440

QY	102	127379	QY	122	127340	QY	140	Db	127292	QY	160	Db	127232	QY	180	Db	127178	QY	200	Db	127121	QY	220	Db	127061	QY	240	Db	127004	QY	259	Db	126950	QY	279	Db	126890	QY	297	Db	126830	QY	314	Db	126785	QY	330	Db	126725	RESULT 14
HisMeth1ShSg1pHeTyrAspSer-----AspSerThrValSerLeu	SerAspHisArgAlaAlaGlnIleArgMetIleGlnIleSerLeuArgPheAlaSerVal	TCTGAT-----TTTGTCATGGAATGCTGCTTTGGGCTGGTTTG	SerGluGluArgSerLeuTrpPro-----LysSerIleValAspValGlyCysGlyIle	-----GATAAACTACCCCTCGTACTACTCTGTGTGATGCTGGTTGGTAAT	GlyGlySerSerArgTyrLeuAlaLysLysPheGlyAlaThrSerValGlyIleThrLeu	GGGGTACTAGTCGCAATTTTGGCAGGGATTTGATGGAATTTGCCGTTACAGTACCAATC	SerProValGlnAlaGlnAlaArgAlaAsnAlaLeuAlaAlaAlaGlnGlyLeuAlaAspLys	AGCCCCCAACAAAGTCCCAAGCGGCTCAGAGTTAACACCAACAGACGATGATGCA-----	ValSerPheGlnValAlaAlaAspAlaLeuGlnGlnIleProPheSerAspGlyGlnPheAspLeu	---CAGTTTTCGTCGAGATGCAATGAGGGGCTTCCTCCCGAATAATAGTTTGTATGTA	ValTrpSerMetGluSerGlyGlyHisMetProAspLysAlaIlePheValGlyLeuLeu	GTTTGCTCAATTTGAAGACGTGGCCCCACATATGCCAGATAAAGCCATTTTGGCCCAAGATTTG	AlaArgValAlaAlaAlaProGlyAlaIleIleIleIleIleValAlaThrTrpCysHisArgAspLeu	ATGCGGCTACTAAAGCCTGTGGATTCATGATCATGTTTAAACCGACTGGAATCAGCAGAC---	GlyProAspGluGlnSer---LeuHisProTrpGluGlnAspLeuLeuLysIleCys	-----GATCGGCCAAAACCCCTCATATTTTGGGAGAAAACGATATGACAGCACTACTA	AspAlaIleTyrTyrLeuProIleTrpCysSerThrSerAspTyrValLysLeuGlnInsSer	GATCAGTGGTCTCATATCCACACTTTTCCACACATGCAAGCGCTTTTCTGAGCTTTTGGCAGCG	LeuSerLeu-----GlnAspIleLysSerGlnAspTrpSerArgPheValAlaProPhe	ACGGATATTAGTAGAAGGGAGGATTAATCACCCGACAGCTGAGCAACAAACACTCCCTCT	TrpProAlaValAlaIleArgSerAlaPheThrTrpLysGlyLeu-----SerSerLeu	TGGCTTGATTCTATC-----TGGCAAGGATATGTTAGACACAGAGATTA	-----LeuSerSerGlyGlnLysThrIleLysGlyAlaLeuAlaLeuLametProLeu	GTGCGTTTGGTCTCATCTGTGTTTCATTAATCTCTGCGGAAGTGGCTTACCTACTACTG	MetIleGlnGlyTyrLysLysAspLeuIleLysPheAlaIle	ATGACGCTGGCATTCGTCACGACGACTCTGATGATTTTGGGATG	AX073663	AX073663	930 bp	DNA	linear	PAT 06-FEB-2001	AX073663	Sequence 7 from Patent WO0104330.	AX073663	AX073663	gi:12710084	AX073663.1	gi:12710084	Synechocystis sp. PCC 6803.	Synechocystis sp. PCC 6803.	Bacteria; Cyanobacteria; Chroococcales; Synechocystis.	1 (bases 1 to 930)	Herbers.K., Badur.R., Kunze.I. and Geisler.M.	Identification and overexpression of a dna sequence coding for	2-methyl-6-phytylhydroquinone-methyltransferase in plants	Patent: WO 0104330-A 7 18-JAN-2001;	Sungene GmbH & Co. KGAA (DB)		

OY 278 SerLeuSerLeuGlnasp-----IleLysSerGluAspTrpSerArgPheValAlaPro 295
 DB 670 GCCACGGGTTGGTGGAGGCCAGGAGTACTGCTGATGAGTACCGACCTCC 729
 OY 296 PheTrpProAlaValIleArgSerAlaPheThrTrpLysGlyLeu 310
 DB 730 GCTTGGTTGGATACCATT-----TGGCAGGGCATT 759

RESULT 15

AX073657

LOCUS

Sequence 1 from Patent WO0104330.

DEFINITION

AX073657

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Source

CDS

1..957

/db_xref="taxon:1148"

/note="unnamed protein product"

/codon_start=1

/transl_table=1

/protein_id="CAC28415.1"

/db_xref="GI:12710079"

/translation="MPEYLILPAGLISLSLAIAAGLYLTARGYSSDVANAYDQMT

EGGLEYMGDHIHGHYGDPPVAKDFOSKIDFVHMAOMGGDLTPGTTVDVGC

GIGSSRIILAKDVGFTNGITISPOQVRAETELPDTAKAVDDAMALSPDSSFD

VYMSVEKGFPHRDKAVFAKELLRYRKPGGILVADPNORDRQVPLNWEKRYAKQLL

DQMSHPAASLEGFALNLEATGLVEGOVTTADVTVPILPAMLDITMGILIRPGWLOT

GIRGFISVREVPILILMLRIALFVGLCRFMKRAVRKNAQA"

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-857-613A-28 (1-350) x AX073657 (1-957)

OY 69 IleAlaGluPheTrpAspGlu-----SerSerGlyLeuTrpGlnAsnIleTrpGlyAsp 86

DB 106 GTGGCGACGCTTACGACCAATGACAGAGAGCGATTTTGGAAATATTACTGGGGCAGC 165

OY 87 HisMetHisHisGlyPheTrpAspSerAspSerThrValSerLeuSerAspHisArgAla 106

DB 166 CATATCCACCTGGCCATTAT---GGCGATCCGCCAGTGGCCAGAT-----210

OY 107 AlaGlnIleArgMetIleGlnIleSerLeuArgPheAlaSerValSerGluArgSer 126

DB 211 -----TTGATCCAAATCGAATAATTGATTGTC-----CATGCCATGGCC 249

OY 127 LysTrp-----ProLysSerIleValAspValGlyGlyGly 138

DB 250 CAGTGGCGGCGATTAGATACACTTCCCGCGACACAGGATATGATGGTGGCGGC 309

OY 139 IleGlyGlySerSerArgTrpLeuAlaLysLysPheGlyValAlaThrSerValGlyIleThr 158

DB 310 ATTGGCGGAGACGATGCGATTCTCGCCAAAGATATGATTGTTTAACGTTACCGGCATCAC 369

OY 159 LeuSerProValGlnAlaGlnArgAlaAsnAlaLeuAlaIleAlaGlnGlyLeuAlaAsp 178
 DB 370 ATTAATCCCAACAGGTGAACAGGGGCGAATTAATCTCTCCGATGTACGGCC---426
 OY 179 LysValSerPheGlnIleAlaAlaAspAlaLeuGlnIleProPheSerAspGlyGlnPheAsp 198
 DB 427 -----AAGTTGGCGGTGAGCATGCTATGCTTGTCTTTCCGACGAGTAGTTGAC 480

OY 199 LeuValTrpSerMetGluSerGlyGluHisMetProAspLysAlaLysPheValGlyLeu 218
 DB 481 GTAGTTTGGTGGTGGAGGAGGCGCCCATGCTGACAAAGCTGTGTGGCCAGGAA 540

OY 219 LeuAlaArgValAlaAlaProGlyValIleIleIleValIleThrTrpCysHisArgAsp 238
 DB 541 TTACTCGCGGTGCGAACCAGGGGCGATCTGCTGCTGGCGGATTTGGAATCACGGGAC 600

OY 239 LeuGlyProAspGluGln---SerLeuHisProTrpTrpGlnIleAspLeuLysLysIle 257
 DB 601 -----GATCGCCAGGTGCCCTCAACTCTGGAAAAACAGTATGCGAAGACTG 651

OY 258 CysAspAlaTrpTrpLeuProAlaTrpCysSerThrSerAspTrpValLysLeuGln 277
 DB 652 TTGATCAATGCTCCACCCCTCTTCCAGCATGAGGTTTTCGCGAAATTTGAA 711

OY 278 SerLeuSerLeuGlnasp-----IleLysSerGluAspTrpSerArgPheValAlaPro 295
 DB 712 GCCACGGGTTGGTGGAGGCCAGGAGTACTGCTGATGAGTACCGACCTCC 771

OY 296 PheTrpProAlaValIleArgSerAlaPheThrTrpLysGlyLeu 310
 DB 772 GCTTGGTTGGATACCATT-----TGGCAGGGCATT 801

Search completed: March 19, 2003, 08:05:54
 Job time: 2726 secs

GenCore version 5.1.4_P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 19, 2003, 05:11:11 : Search time 221 Seconds
(without alignments)

3566.513 Million cell updates/sec

Title: US-09-857-613a-28

Perfect score: 1830

Sequence: 1 MATVVRIPRTISCIHHTFRS.....IEGYKKDLIKFAITCKRPE 350

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV-xlh
-MODE=frame_p2n_model -DEV-xlh
-O=/gcn2.1/USPTO.spool/US09857613/runat.12032003.100035.6003/app.query.fasta_1.519
-DB=N.Geneseq.101002 -OFMT=fastap -SUFFIX=png -MINMATCH=0.1 -LOOPL=0
-DOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=DIOSum62 -TRANS=human60.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFM=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09857613.ecgn.1.1.200.0runat.12032003.100035.6003 -NCPU=6 -ICPU=3
-NO_XIPYX -NO_MMAP -LARGEOUSER -NEG_SCORES=0 -WAIT -LONGCIS -DEV_TIMECUT=120
-WARN_TIMECUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-TGAPOP=10 -TGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: N.Geneseq.101002.*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
25: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1830	100.0	1189	21	AAA29164	Soybean gamma-toco
2	1668.5	91.2	1331	21	AAA29154	Soybean gamma toco
3	1173	64.1	1257	21	AAA29165	Soybean gamma-toco
4	1162	63.5	1102	21	AAA29161	Corn gamma-tocophe
5	1154.5	63.1	1293	21	AA050535	Arabidopsis thalia
6	1151.5	62.9	1071	21	AA298317	A. thaliana gene 1
7	1151.5	62.9	1350	24	AA46039	A thaliana gamma-t
8	1151.5	62.9	1790	20	AA417789	Arabidopsis gamma-t
9	1147.5	62.7	1296	21	AA034022	Arabidopsis thalia
10	1022.5	55.9	1862	22	AAH44261	Physcomitrella pat
11	956	52.2	1011	21	AAA29155	Wheat gamma tocophe
12	778	42.5	792	21	AAA29151	Corn gamma tocophe
13	647.5	35.4	954	20	AA417788	Synechocystis gamm
14	392	21.4	930	22	AA026173	Synechocystis PCC6
15	392	21.4	957	21	AA261599	DNA encoding a met
16	392	21.4	957	22	AA026169	Synechocystis PCC6
17	392	21.4	957	24	AA46040	Synechocystis 2-me
18	392	21.4	974	22	AA026170	Synechocystis PCC6
19	344.5	18.8	521	21	AAA29152	Partial gamma toco
20	344.5	18.8	521	21	AAA29162	Rice gamma-tocophe
21	316	17.3	311	24	AB74098	Corn tassels-deriva
22	287	15.7	53500	21	AA055842	Complete nucleotid
23	280.5	15.3	852	21	AA055802	S. lavenulae M1M
24	267.5	14.6	464	21	AA029163	Rice gamma-tocophe
25	265	14.5	488	22	AAH44245	Physcomitrella pat
26	261	14.3	2122	18	AA070152	S. longisporoflavus
27	236	12.9	1410	15	AA080524	Oxidoreducting aver
28	232.5	12.7	2814	21	AA050960	A. halophila betai
29	229.5	12.5	592	21	AAA29153	Rice gamma tocophe
30	227	12.4	1383	21	AA029164	Maize C-24 sterol
31	226	12.3	1149	21	AA044838	Arabidopsis thalia
32	226	12.3	3219	21	AA050961	E. halochloris bet
33	225	12.3	6085	18	AA070153	S. longisporoflavus
34	224.5	12.3	1497	22	AA070839	Corn SMT cDNA. 2e
35	220.5	12.0	828	22	AA088318	S. spinosa DNA fra
36	220.5	12.0	45624	22	AA088315	S. spinosa DNA fra
37	220.5	12.0	50000	22	AA088312	S. spinosa DNA fra
38	220.5	12.0	80161	20	AA021501	DNA fragment of Sa
39	218	11.9	12381	21	AA058381	Streptomyces averm
40	207.5	11.3	1041	24	AA03672	Nicotiana tabacum
41	204.5	11.2	1233	21	AA048513	Arabidopsis thalia
42	203.5	11.1	1235	21	AA039342	Arabidopsis thalia
43	203.5	11.1	1270	21	AA048512	Arabidopsis thalia
44	203.5	11.1	1274	21	AA039297	Arabidopsis thalia
45	203.5	11.1	1320	20	AA070837	Yeast SMT gene. S

ALIGNMENTS

RESULT 1
AAA29164
ID AAA29164 standard; cDNA: 1189 BP.

AC AAA29164;

DT 12-SEP-2000 (first entry)

XX Soybean gamma-tocopherol methyltransferase cDNA (clone sah1c.pk004.g2).

DE Vitamin E: alpha-tocopherol; biosynthesis; enzyme; inhibitor; herbicide;

KW gamma-tocopherol methyltransferase; ss.

XX Glycine max.

OS Key Location/Qualifiers
FH CDS 5..1057
FT /*tag= a

/product- gamma-tocopherol_methyltransferase

FT XX
 PN W0200032757-A2.
 PD 08-JUN-2000.
 XX
 PF 02-DEC-1999; 99WO-US28588.
 XX
 PR 03-DEC-1998; 98US-0110781.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Cahoon RE, Coughlan SJ, Miao G, Rafalski JA;
 XX
 DR WPI: 2000-412309/35.
 XX P-PSDB: AAY96472.
 XX
 PT Polynucleotide encoding a vitamin E biosynthetic enzyme polypeptide
 PT such as gamma tocopherol methyltransferase, useful for synthesizing
 PT vitamin E or as a target for design and discovery of herbicides
 XX
 PS Claim 3; Page 64-65; 82pp: English.
 XX
 CC AAA29151-69 encode vitamin E (alpha-tocopherol) biosynthetic enzymes or
 CC their fragments. The cDNA clones were identified by BLAST searches based
 CC on similarity to Synecocystis sp. and Arabidopsis thaliana vitamin E
 CC biosynthetic enzymes. The enzymes are useful for synthesizing
 CC plastoquinones or tocopherols (especially vitamin E). Vitamin E is
 CC required for plant growth, therefore the enzymes may be used for the
 CC discovery of new herbicides. The enzymes can be used in methods to
 CC evaluate potential inhibitors, which may have use as herbicides.
 CC Additionally, transgenic expression of, e.g. gamma-tocopherol
 CC methyltransferase affords the ability to manipulate tocopherol levels as
 CC desired for a particular application.
 XX
 SQ Sequence 1189 BP; 333 A; 253 C; 299 G; 304 T; 0 other:
 XX
 Alignment Scores:
 Pred. No.: 2,34e-180 Length: 1189
 Score: 1830.00 Matches: 350
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-857-613a-28 (1-350) x AAA29164 (1-1189)

QY 1 MetAlaThrValAlaArgIleProThrIleSerCysIleHisIleHisThrPheArgSer 20
 DB 5 ATGGCCACCGGTGGAGATGCCAACAATCATGATGCCATCCACACGCTTCCGTTCC 64
 QY 21 GluSerProArgThrPheAlaArgIleArgValGlyProArgSerTTPAlaProIleArg 40
 DB 65 CAATCCCTCCGCTTCGCCAGAAATCCGGGTCCGACCCAGAGCTGTCGCTTATTCGG 124
 QY 41 AlaSerIleAlaSerSerGluArgGlyGluIleValleuGluGlnIlyProLysLysasp 60
 DB 125 GCATCGCGAGGAGCTGGAGAGAGGAGATGATTGAGCGCAAGAACCCAAAGAGAT 184
 QY 61 AspLysLysLysLeuGlnLysGlyIleAlaGluPheTyrAspGluSerSerGlyLeuTyrP 80
 DB 185 GACAGAAGAAGATCGCAGAAGGAAATCGCAAGATTTTACGACGATGCTTCCGCTTATGG 244
 QY 81 GluAsnIleTrrpGlyAspHisMetHisGlyPheTyrAspSerAspSerThrValSer 100
 DB 245 GAGAACATTTTGGGCGACACATGACCATGCTTTTATGATCGATTCACATGTTTCG 304
 QY 101 LeuSerAspHisArgAlaAlaGlnIleArgMetIleGlnIlySerIleArgPheAlaSer 120
 DB 305 CTTTCGATCATCTGCTGCTCAATCGAATGCCAAGATCTCTTCGCTTTCCTCT 364
 QY 121 ValSerGluGluArgSerLysTrrpProLysSerIleValAlaValGlyCysGlyIleGly 140

DB 365 GTTTCGAGAGACCTAGTAATAGGCCCAAGAGATATAGTTGATGTGGGTGTGCATAGT 424
 QY 141 GlySerSerArgTrrpIleAlaAlaLysLysPheGlyAlaThrSerValGlyIleThrLeuSer 160
 DB 425 GGCAGCTCTGATACCTGGCCAAAGAAATTTGGACCAACAGATGAGCATACTCTGAGT 484
 QY 161 ProValGlnAlaGlnArgAlaAsnAlaLeuAlaAlaAlaGlnIlyLeuAlaAspLysVal 180
 DB 485 CTTGTTCAAGCTCAAGAGCAAAATGCTCTCTGCTGCTCAAGATTTGGCTGATAGGTT 544
 QY 181 SerPheGlnValAlaAlaAspAlaLeuGlnIleProPheSerAspGlyGlnPheAspLeuVal 200
 DB 545 TCTTTTCAGGTGTGAGAGCTCTACAGAACCATTTCTTGACGGCAGCATTTGATCTGCTG 604
 QY 201 TrrpSerMetGluSerGlyGlnHisMetProAspLysAlaLysPheValGlyGluLeuAla 220
 DB 605 TGGTCCATGGAGATGGAGACATATGCTTCACAAAGCTAAGTTGTTGGAGAGTTACT 664
 QY 221 ArgValAlaAlaProGlyAlaIleIleIleIleValThrTrrpCysHisArgAspLeuGly 240
 DB 665 CGGTAGCAGCACCAAGGTGCCATTTATATATAGTACATGCTGCCACAGGAGATCTTGGC 724
 QY 241 ProAspGluGlnSerLeuHisProTrrpGluGlnAspLeuLysLysIleCysAspAla 260
 DB 725 CCTGACGACAAATCCTTACATCCATGGGAGCAAGATCTTAAAGAAATTTGGCATGCA 784
 QY 261 TyrTrrpLeuProAlaIleTrrpCysSerThrSerAspIlyValLysLeuGlnIleSerLeuSer 280
 DB 785 TATTACTCTCCCTGGCTGCTCAACTGTATATATGTTAAGTTGCTCAATCCCTGTCA 844
 QY 281 LeuGlnAspIleLysSerGluAspTrrpSerArgPheValAlaProPheTrrpProAlaVal 300
 DB 845 CTTAGACATCAAGATCGAAGAGATTTGGCTGCTTGTGCTCCATTTTGGCCAGCAGTG 904
 QY 301 IleArgSerAlaPheThrTrrpLysGlyLeuSerSerLeuLeuSerSerGlyGlnLysThr 320
 DB 905 ATACGCTACGCTTCACATGAGAGGCTATCTTCACTGCTGAGACAGCAAAAGAAC 964
 QY 321 IleLysGlyAlaLeuAlaMetProLeuMetIleGluGlyTrrpLysLysAspLeuIleLys 340
 DB 965 ATAAAGAGAGCTTGGCTATGCTCATTTGATGATAGAGGATACAGAAAGATCTAATTAAAG 1024
 QY 341 PheAlaIleIleThrCysArgLysProGlu 350
 DB 1025 TTTGCCATCATTCATGTCGAAACCTGAA 1054

RESULT 2
 AAA29154
 ID AAA29154 standard; cDNA; 1331 BP.
 XX
 AC AAA29154:
 XX
 AC 12-SEP-2000 (first entry)
 XX
 DE Soybean gamma tocopherol methyltransferase contig cDNA.
 KM Vitamin E: alpha-tocopherol; biosynthesis; enzyme; inhibitor; herbicide;
 KM gamma-tocopherol methyltransferase; ss.
 XX
 OS Glycine max.
 XX
 FH key Location/Qualifiers
 FT CDS 5..1055
 FT /*tag= a
 FT /product= gamma_tocopherol_methyltransferase
 PN W0200032757-A2.
 XX
 PD 08-JUN-2000.
 XX
 PF 02-DEC-1999; 99WO-US28588.
 XX
 PR 03-DEC-1998; 98US-0110781.

XX (DUPO) DU PONT DE NEMOURS & CO E I.
 PA
 XX
 XX Cahoan RE, Coughlan SJ, Miao G, Rafalski JA:
 XX WPI: 2000-412309/35.
 DR P-PSDB: AAY96462.
 XX
 XX Polynucleotide encoding a vitamin E biosynthetic enzyme polypeptide
 PT sequence encoding a vitamin E methyltransferase, useful for synthesizing
 PT vitamin E or as a target for design and discovery of herbicides
 PT
 XX
 PS Claim 3: Page 52: 82pp: English.
 PS
 XX AAA29151-69 encode vitamin E (alpha-tocopherol) biosynthetic enzymes or
 CC their fragments. The cDNA clones were identified by BLAST searches based
 CC on similarity to Synecocystis sp. and Arabidopsis thaliana vitamin E
 CC biosynthetic enzymes. The enzymes are useful for synthesizing
 CC plastoquinones or tocopherols (especially vitamin E). Vitamin E is
 CC required for plant growth, therefore the enzymes may be used for the
 CC discovery of new herbicides. The enzymes can be used in methods to
 CC evaluate potential inhibitors, which may have use as herbicides.
 CC Additionally, transgenic expression of, e.g., gamma-tocopherol
 CC methyltransferase affords the ability to manipulate tocopherol levels as
 CC desired for a particular application.
 CC
 XX
 XX Sequence 1331 BP: 373 A; 293 C; 301 G; 364 T; 0 other:

Alignment Scores:

Pred. No.: 1,666-163 Length: 1331
 Score: 1668.50 Matches: 323
 Percent Similarity: 94.52% Conservative: 5
 Best Local Similarity: 93.08% Mismatches: 10
 Query Match: 91.17% Indels: 9
 DB: 21 Gaps: 1

US-09-857-613a-28 (1-350) x AAA29154 (1-1331)

OY 1 MetAlaThrValAlaArgIleProThrIleSerCysIleHisIleHisThrPheArgSer 20
 DB 6 ATGGCCACCGTGGTGAAGATCCCAACATCTCATGTCACATCCACAGTCCGCTTC 65
 OY 21 GlnSerProArgThrPheAlaArgIleArgValGlyProArgSerTrpAlaProIleArg 40
 DB 66 CAATCCCTCGCAGCTTCCGCCAATCCGGGTGGACCCAGGCTGGGCTCTATTCCG 125
 OY 41 AlaSerAlaIleSerSerGluArgGlyIleValIleGluIleGlnIleProIleValAsp 60
 DB 126 GCATTCGGCAGCGCTGGAGAGGGGAGATGATGAGGAGCAAGCCGAGAGAGGAT 185
 OY 61 AspLysLysLysLeuGlnIleGlyIleAlaGluPheTrpAspGlnSerSerGlyLeuTrp 80
 DB 186 GACAAAGAAAGCTGCAGAAAGGAATCGCAGATTTTACGACAGTCGTGGCTTATGG 245
 OY 81 GluAsnIleTrpGlyAspHisMetHisGlyPheTrpAspSerAspSerThrValSer 100
 DB 246 GACAAATTTGGGGCCAGCAGATGACATGCTTTATGATCGATTCACACTGTTTCG 305
 OY 101 LeuSerAspHisArgAlaAlaGlnIleArgMetIleGlnIleSerLeuArgPheAlaSer 120
 DB 306 CTTTCGATCATCGCTGCTCGATCCGATCCGAATATCCAAAGACTCTTCGCTTGCCTCT 365
 OY 121 ValSerGluGlnArgSerLysTrpProLysSerIleValAspValGlyCysGlyIleGly 140
 DB 366 GTTTCGAGAGAGCTGAATATGCCCCAGAGATGATGTTGGGTGGCATAGGT 425
 OY 141 GlySerSerArgTrpLeuAlaLysLysPheGlyAlaThrSerValGlyIleThrLeuSer 160
 DB 426 GCGAGCTGATACCTGCGCAAGAAATTTGAGCAACAGTGAAGCATCAGCTCAGT 485
 OY 161 ProValGlnAlaGlnArgAlaAsnAlaLeuAlaAlaAlaGlnIleGlyLeuAlaAspLysVal 180
 DB 486 CCGTTCAACCTCAAAAGCAAAATGCTTGTGCTGCTCAAGGATTGGCTGATTAAGCTT 545

OY 181 SerPheGlnValAlaAspAlaLeuGlnIleProPheSerAspGlyGlnPheAspLeuVal 200
 DB 546 TCCCTTCAGGTTGCTGACGCTCTACAGCAACCATCTCTGACGGCCAGTTGATCTGGTG 605
 OY 201 TrpSerMetGluSerGlyGluHisMetProAspLysAlaLysPheValGlyGluLeuAla 220
 DB 606 TGGTCATGAGAGAGTGCAGAGCATATGCTGCACAAAGCTAAGTTGTTGGAGAGTTACT 665
 OY 221 ArgValAlaAlaProGlyAlaIleIleIleValIleThrProCysHisArgAspLeuGly 240
 DB 666 CGGGTAGAGAGCAGCAGGTCCTATTAATATAGTAACATGGTGCACAGGATCTTGCC 725
 OY 241 ProAspGluGlnSerLeuHisProTrpGluGlnAspLeuLeuLysLysIleCysAspAla 260
 DB 726 CCGAGCAACATCTTATATCATGCGGAGCAAGATCTCTTAAAGAAATTTGGCATTCGA 785
 OY 261 TyrTrpLeuProAlaIleTrpCysSerThrSerAspTrpValLysLeuGlnSerLeuSer 280
 DB 786 TATTACCTCCCTGCTGCTGCTCAACTTGTGATTTATGTTACTTGCATCCATCCCTGTCA 845
 OY 281 LeuGlnAspIleLysSerGluAspTrpSerArgPheValAlaProPheTrpProAlaVal 300
 DB 846 CTTCAAGACATCAAGTACAGACATTTGCTCGCTTGTGCTGCATTTTGGCCAGCAGTG 905
 OY 301 IleArgSerAlaPheThrTrpLysGlyLeuSerSerLeuLeuSerSerGlyGlnIleThr 320
 DB 906 ATACGCTCAGCTTCATCATGAGGAGGCTTATCTTCACTTGAAGCAGTGAAGCTTGA 965
 OY 321 IleLysGlyAlaLeuAlaMetProLeuMetIleGlyIleLysLysAspLeuIleLys 340
 DB 966 ATTATATATGCA-----TTTCAAAACAAACCCGCCCA 998
 OY 341 PheAlaIleIlePheCysArg 347
 DB 999 TCTTATATGCAACTGCAAG 1019

RESULT 3
 AAA29165
 ID AAA29165 standard; cDNA; 1257 BP.
 XX
 AC AAA29165;
 XX
 DT 12-SEP-2000 (first entry)
 XX
 DE Soybean gamma-tocopherol methyltransferase cDNA.
 XX
 KW Vitamin E: alpha-tocopherol; biosynthesis; enzyme; inhibitor; herbicide;
 KW gamma-tocopherol methyltransferase; ss.
 XX
 OS Trilicium aestivum.
 XX
 FH Key Location/Qualifiers
 FT CDS 18..1118
 FT /tag= a
 FT /transl_except= (pos:30..32, aa:Xaa)
 FT /transl_except= (pos:150..152, aa:Xaa)
 FT /product= gamma-tocopherol methyltransferase
 FT /note= "Xaa is not defined"
 PN
 XX
 WO200032757-A2.
 PD 08-JUN-2000.
 XX
 PF 02-DEC-1999: 99NO-US28588.
 XX
 PR 03-DEC-1998: 98US-0110781.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Cahoan RE, Coughlan SJ, Miao G, Rafalski JA:
 DR WPI: 2000-412309/35.

QY 233 ThrTrpCysHisArgAspLeuGlyProAspGlnGlnSerLeuHisProTrpGlnGlnAsp 255

required for plant
discovery of new herb

AAA23151-66 encode vitamin E (alpha-tocopherol) biosynthetic enzymes or their fragments. The cDNA clones were identified by BLAST searches based on similarity to *Synechocystis* sp. and *Arabidopsis thaliana* vitamin E biosynthetic enzymes. The enzymes are useful for synthesizing plastoquinones or tocopherols (especially vitamin E). Vitamin E is required for plant growth, therefore the enzymes may be used for the recovery of new herbicides. The enzymes can be used in methods to

CC evaluate potential inhibitors, which may have use as herbicides.
CC Additionally, transgenic expression of, e.g. gamma-tocopherol
CC methyltransferase affords the ability to manipulate tocopherol levels as
CC desired for a particular application.

XX Sequence 1102 BP; 231 A; 330 C; 328 G; 207 T; 6 other:

Alignment Scores:

Pred. No.:	5.5e-111	Length:	1102
Score:	1162.00	Matches:	225
Percent Similarity:	75.88%	Conservative:	33
Best Local Similarity:	66.18%	Mismatches:	72
Query Match:	63.50%	Indels:	10
DB:	21	Gaps:	2

US-09-857-613A-28 (1-350) x AAA29161 (1-1102)

```
OY 20 SerGlnSerProArgThrPheAlaArgValGlyProArgSerTrpAlaProIle 39
    |||||
DB 28 TCCTGAGCTCCAGAGAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAG 87
OY 40 ArgAlaSerAlaAlaSerSerGluArgGlyGluIleValLeuGlnLeuIleProlLys 59
    |||
DB 88 CACGTCCCGCGGCGCACTCCCGCGCTCCGACGCGCGCTGTCACGCTGGCGATGGCC 147
OY 60 AspAspLysLysLys-----LeuGlnLysGlyIleAlaGlu 71
    |||
DB 148 TCGTCGACGCGCTCAGCGCCCGCGCGCGCGCGCTCTGTAAGAGAGGCACTCGCGGG 207
OY 72 PheTrpAspGlnSerSerGlyLeuTrpGluAsnIleTrGlyAlaPheIleHisGly 91
    |||||
DB 268 TCCTACGAGCTCCAGAGAGCGCCCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCT 327
OY 112 IleGlnIleSerLeuArgPheAlaSerVal-----SerGluLysArgSerLysTrpPro 129
    |||||
DB 338 ATCGAGGAGCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCT 387
OY 130 LysSerIleValAlaGlnGlyCysGlyIleGlyIleSerSerArgTrpLeuAlaLys 149
    |||||
DB 388 AAACATATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGAT 447
OY 150 PheGlyAlaThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaAsnAla 169
    |||||
DB 448 TACGAGAACCCAGTGCACCTGCGATCAGCTTGAGCCCTTCCTCAACCCGAGAGAGAAATGCT 507
OY 170 LeuAlaAlaAlaGlnGlyLeuAlaAspLysValSerPheGlnValAlaAlaAspAlaLeuGln 189
    |||||
DB 508 CTCGCTGACGCGAGGAGGTTGTCGATCAGTACTCTCAAGTCTGCTGCTGAG 567
OY 190 GlnProPheSerAspGlyLysIlePheAspLeuValTrpSerMetCysLeuGlyIleHisMet 209
    |||||
DB 568 CAACGCTTCCCTGACGCGAGGATTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 627
OY 210 ProAspLysAlaLysPheValGlyGluLeuAlaArgValAlaAlaProGlyAlaIleIle 229
    |||||
DB 628 CCGGACAGAGAAAGTTTCTTACTGAGCTAGACCGGCTGCGCTCTCTGAGGAGGACATA 687
OY 230 IleIleValThrTrpCysHisArgAspLeuGlyProAspGlnLeuSerLeuHisProTrp 249
    |||||
DB 688 ATCATGCTGACATGCTGATGAGAACCTGATGATCCGAAACCTGCTAAAGCCGAT 747
OY 250 GluGlnAspLeuLeuLysLysIleCysAspAlaTyrTrpLeuProAlaTrpCysSerThr 269
    |||||
DB 748 GAACCTGACCTCTCTGAGGAGATATCCAGCGCTACTACCTCCGAGCTGCTCTCCCT 807
OY 270 SerAspTrpValLysLeuGlnSerLeuSerLeuGlnAspIleLysSerGluAspTrp 289
    |||||
DB 808 TCAGATATGTCGAACATTCGCAAGTCACTCTCTCTGAGAGATATCAAGACAGCTGACTGG 867
```

```
OY 290 SerArgPheValAlaProPheTrpProAlaValIleArgSerAlaPheThrTrpLysGly 309
    |||
DB 868 TCAGGAGAACGTGCGCCGCTTTGGCCCGCTGATAAATCAGCCCTAACATGAGAGGC 927
OY 310 LeuSerSerLeuLeuSerSerGlyLysTrpIleLysGlyAlaLeuAlaMetProLeu 329
    |||||
DB 928 TTCACCTCTCTCTGACGACCGGATGAGAACAGATCAGAGGCGCATGTGATGCCGCTA 987
OY 330 MetIleGlyIleTrpLysLysAspLeuIleLysPheAlaIleIleThrCysArgLysPro 349
    |||||
DB 988 ATGATCCAGGCTACAGAAAGGCGCTCATCAATTCACCATCATCAGCTGTGCAAGCCT 1047
RESULT 5
AAC50535
ID AAC50535 standard; DNA; 1293 BP.
AC AAC50535;
DE 18-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 65205.
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 65205.
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
```

PR	08-JUN-1999	9905-0138094
PR	10-JUN-1999	9905-0138540
PR	10-JUN-1999	9905-0138847
PR	14-JUN-1999	9905-0139119
PR	16-JUN-1999	9905-0139452
PR	16-JUN-1999	9905-0139453
PR	17-JUN-1999	9905-0139452
PR	18-JUN-1999	9905-0139454
PR	18-JUN-1999	9905-0139455
PR	18-JUN-1999	9905-0139456
PR	18-JUN-1999	9905-0139463
PR	18-JUN-1999	9905-0139465
PR	18-JUN-1999	9905-0139463
PR	18-JUN-1999	9905-0139750
PR	21-JUN-1999	9905-0139763
PR	22-JUN-1999	9905-0139817
PR	22-JUN-1999	9905-0139839
PR	23-JUN-1999	9905-0140035
PR	23-JUN-1999	9905-0140354
PR	24-JUN-1999	9905-0140635
PR	28-JUN-1999	9905-0140823
PR	30-JUN-1999	9905-0140991
PR	30-JUN-1999	9905-0141287
PR	01-JUL-1999	9905-0141842
PR	01-JUL-1999	9905-0142154
PR	02-JUL-1999	9905-0142055
PR	06-JUL-1999	9905-0142390
PR	08-JUL-1999	9905-0142803
PR	09-JUL-1999	9905-0142920
PR	12-JUL-1999	9905-0142977
PR	13-JUL-1999	9905-0143542
PR	14-JUL-1999	9905-0143624
PR	15-JUL-1999	9905-0144005
PR	16-JUL-1999	9905-0144085
PR	16-JUL-1999	9905-0144086
PR	19-JUL-1999	9905-0144325
PR	19-JUL-1999	9905-0144331
PR	19-JUL-1999	9905-0144332
PR	19-JUL-1999	9905-0144333
PR	21-JUL-1999	9905-0145086
PR	21-JUL-1999	9905-0145088
PR	22-JUL-1999	9905-0145085
PR	22-JUL-1999	9905-0145087
PR	22-JUL-1999	9905-0145089
PR	23-JUL-1999	9905-0145145
PR	23-JUL-1999	9905-0145218
PR	23-JUL-1999	9905-0145224
PR	26-JUL-1999	9905-0145276
PR	27-JUL-1999	9905-0145913
PR	27-JUL-1999	9905-0145918
PR	27-JUL-1999	9905-0145919
PR	28-JUL-1999	9905-0145951
PR	02-AUG-1999	9905-0146386
PR	02-AUG-1999	9905-0146388
PR	02-AUG-1999	9905-0146389
PR	03-AUG-1999	9905-0147038
PR	04-AUG-1999	9905-0147204
PR	04-AUG-1999	9905-0147302
PR	05-AUG-1999	9905-0147192
PR	05-AUG-1999	9905-0147260
PR	06-AUG-1999	9905-0147303
PR	06-AUG-1999	9905-0147416

PR	09-AUG-1999;	9905-0147493;
PR	09-AUG-1999;	9905-047935;
PR	10-AUG-1999;	9905-0148171;
PR	11-AUG-1999;	9905-0148319;
PR	12-AUG-1999;	9905-0148341;
PR	13-AUG-1999;	9905-0148565;
PR	13-AUG-1999;	9905-0148684;
PR	16-AUG-1999;	9905-0149368;
PR	17-AUG-1999;	9905-0149175;
PR	18-AUG-1999;	9905-0149426;
PR	20-AUG-1999;	9905-0149723;
PR	20-AUG-1999;	9905-0149722;
PR	20-AUG-1999;	9905-0149929;
PR	23-AUG-1999;	9905-0149902;
PR	23-AUG-1999;	9905-0149930;
PR	25-AUG-1999;	9905-0150566;
PR	26-AUG-1999;	9905-0150884;
PR	27-AUG-1999;	9905-0151065;
PR	27-AUG-1999;	9905-0151066;
PR	27-AUG-1999;	9905-0151080;
PR	30-AUG-1999;	9905-0151300;
PR	31-AUG-1999;	9905-0151438;
PR	01-SEP-1999;	9905-0151930;
PR	07-SEP-1999;	9905-0152363;
PR	10-SEP-1999;	9905-0153070;
PR	13-SEP-1999;	9905-0153758;
PR	15-SEP-1999;	9905-0154018;
PR	16-SEP-1999;	9905-0154039;
PR	20-SEP-1999;	9905-0154779;
PR	22-SEP-1999;	9905-0155139;
PR	23-SEP-1999;	9905-0155486;
PR	24-SEP-1999;	9905-0155659;
PR	28-SEP-1999;	9905-01556458;
PR	29-SEP-1999;	9905-0155566;
PR	04-OCT-1999;	9905-0157117;
PR	05-OCT-1999;	9905-0157753;
PR	06-OCT-1999;	9905-0157865;
PR	07-OCT-1999;	9905-0158029;
PR	08-OCT-1999;	9905-0158232;
PR	12-OCT-1999;	9905-0158533;
PR	13-OCT-1999;	9905-0159293;
PR	13-OCT-1999;	9905-0159294;
PR	13-OCT-1999;	9905-0159329;
PR	13-OCT-1999;	9905-0159325;
PR	14-OCT-1999;	9905-0159330;
PR	14-OCT-1999;	9905-0159337;
PR	14-OCT-1999;	9905-01596371;
PR	18-OCT-1999;	9905-0159584;
PR	21-OCT-1999;	9905-0160741;
PR	21-OCT-1999;	9905-0160767;
PR	21-OCT-1999;	9905-0160768;
PR	21-OCT-1999;	9905-0160770;
PR	21-OCT-1999;	9905-0160814;
PR	21-OCT-1999;	9905-0160850;
PR	22-OCT-1999;	9905-0160980;
PR	22-OCT-1999;	9905-0160981;
PR	22-OCT-1999;	9905-0160989;
PR	25-OCT-1999;	9905-0161404;
PR	25-OCT-1999;	9905-0161405;
PR	25-OCT-1999;	9905-0161406;
PR	26-OCT-1999;	9905-0161359;
PR	26-OCT-1999;	9905-0161360;
PR	26-OCT-1999;	9905-0161361;
PR	28-OCT-1999;	9905-0161920;
PR	28-OCT-1999;	9905-0161922;
PR	28-OCT-1999;	9905-0161993;
PR	29-OCT-1999;	9905-0162142;

Alignment Scores:	4.15e-110	length:	1293
Pred. No.:		Matches:	231
Scores:	1154.50	Conservative:	30
Percent Similarity:	75.43%		

Best Local Similarity: 66.76% Mismatches: 65
 Query Match: 63.09% Indels: 21
 DB: 21 Gaps: 4

US-09-857-613a-28 (1-350) x AAC50535 (1-1293)

QY 8 Prothrillesercysllehisllehisrhpheargserglnserproargthrphela 27
 |||||
 DB 118 CCAACTCTCTTTCGGCTCAAGTCATCGCTTCTTT-CGGTCTCCATCCTCTCTCC 176
 QY 28 Argileargval---GlyProargsertrpalaProillearglaseralalaser 46
 |||||
 DB 177 TCACTCTCTATGACGACACCGCTGGAACGCTGTGGCGGCTGCTGCTACATCCACT 236
 QY 47 GluarggluilevalleuglnlnsProlysAspaspLysLysleugln 66
 |||||
 DB 237 GAG-----GCGCTAAGA 248
 QY 67 LysglyliealagluPhetyrAspGlnserSerGlyLeutrpGluAsnIletrpGlyasp 86
 |||||
 DB 249 AAAGAAATGCGGAGATTCTACAACTCGGCTTGTGGAGAGATTGGGAGAT 308
 QY 87 HismethisHISglYPhetyrAspserAspserThrValserLeuSerAsp-----His 104
 |||||
 DB 309 CATATGCTCATGCTTATGACCCCTGATCTTCTGTTCACTTCTGATCTGCTCAC 368
 QY 105 ArgAlaalaaglnleargMetlleGlnlnserLeuArgpheaIserValser---Glu 123
 |||||
 DB 369 AAGGAAGCTCAGATCCGATGATGAGAGCTCTCCGTTTGGCGGTACTGATGAA 428
 QY 124 GluargserlystrProlysSerlleValaspValGlyCysGlyTlleGlyLysSer 143
 |||||
 DB 429 GAGGAGGAGAAAGATAAAGAAAGTACTGATGCTGCTGGATGGAGAGAGCTCA 488
 QY 144 ArgTyrleualLysLysPheglValarThrSerValGlylleThrleuSerProvalGln 163
 |||||
 DB 489 AGATATCTTCCCTCTAAATTGAGACCTGATGCTGATGCTGATGCTGATGCTGATG 548
 QY 164 AlaGlnArgAlaAsnAlaLeuAlaAlaGlnGlyLeuAlaAspLysValserPheGln 183
 |||||
 DB 549 GCCAAGAGAGCCAAATGATCTCGCGCTGCTCAATCAGCTGCTCAATGAGCTTCCCA 608
 QY 184 ValAlaAspAlaLeuGlnGlnProPheSerAspGlyGlnPheAspLeuValtrPserMet 203
 |||||
 DB 609 GTTGGGATGCTTGGATGACGACCTTCGAAGATGAAATTCGATCTGCTGCTGAG 668
 QY 204 GlnserGlyLysHisMetProAspLysAlaLysPheValGlyLeuAlaArgValAla 223
 |||||
 DB 669 GAGAGTGTGACATATGCTGACAGGCCAAGTTGTAAAGATGTGCTGCTGCTGCTG 728
 QY 224 AlaProGlyAlaIlellellelleValThrTrpCysHisArgAspLeuGlyProAspGlu 243
 |||||
 DB 729 GCTCCAGAGAGGATGATATATATGATGATGCTGCTGCTGCTGCTGCTGCTGCTG 788
 QY 244 GlnserLeuHisProTrpGlnGlnAspLeuLysLysLysLysAspAlaTyrTrleu 263
 |||||
 DB 789 GAAGCTTTGACACCTGGGAGGCAAAACATCTTGACAAATCTGTAAAGACGTTCTATCTC 848
 QY 264 ProAlaTrpCysSerThrSerAspTyrValLysLeuGlnGlnSerLeuSerLeuGlnAsp 283
 |||||
 DB 849 CCGGCTGTGCTCCACCGATGATGATGATGATGATGATGATGATGATGATGATGATG 908
 QY 284 IleLysSerGluAspTrpSerArgPheValAlaProPheTrpProAlaValIleArgSer 303
 |||||
 DB 909 ATTAAGTGTGGGATTTGGTCAGAGAAAGTACGCTCTTCTGCGCTCGGTTATACGACT 968
 QY 304 AlaPheThrTrpLysGlyLysSerSerLeuLeuSerSerGlyGlnLysThrIleLysGly 323
 |||||
 DB 969 GCATTAACTGAGAGGCGCTTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1028
 QY 324 AlaIleAlaMetProleuMetIleGlnGlyTyrLysLysAspLeuLysPheAlaIle 343
 |||||
 DB 1029 GCATTACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1088

QY 344 IleThrCysArgLysPro 349
 |||||
 DB 1089 ATCACTTCCAGAACCA 1106

RESULT 6

AA298317
 ID AA298317 standard; DNA: 1071 BP.

AA298317;
 14-JUN-2000 (first entry)

DE A. thaliana gene involved in environmental stress tolerance.

DE Environmental stress; plant; transgenic plant; anaerobic; flooding; cold;

DE dehydration; drought; heat stress; salinity; osmotolerance; ds.

OS Arabidopsis thaliana.

PN WO200008187-A2.

PD 17-FEB-2000.

PE 04-AUG-1999; 99WO-EP05652.

PR 04-AUG-1998; 98EP-0202634.

PA (VLAAs-) VLAAMS INTERNUNIVERSITAIR INST BIOTECHNOG.

PI Lee JH, Verbruggen N;

DR WPI: 2000-205726/18.

DR P-PSDB; AAV77937.

PT Isolation of polynucleic acids useful for producing transgenic plant by

isolating genes involved in tolerance to environmental stress

PS Claim 4; Page 130-132; 312pp; English.

CC The invention relates to isolation of coding sequences and/or genes

CC involved in tolerance to environmental stress in plants. The sequences

CC (AA98305-298365) are useful for producing a transgenic plant having

CC enhanced tolerance or resistance to environmental stress conditions such

CC as anaerobic, flooding, cold, dehydration, drought, heat stress or

CC salinity. This is useful for producing improved yield, growth,

CC development and productivity under environmental stress conditions, and

CC also provides growth of crops in areas where they cannot grow without

CC the induced osmotolerance. Sequences AA298305-365 represent

CC polynucleotide sequences from A. thaliana that are involved in

CC environmental stress tolerance.

SQ Sequence 1071 BP; 277 A; 229 C; 270 G; 295 T; 0 other;

Alignment Scores:
 Pred. No.: 6,52e-110 Length: 1071

Score: 1151.50 Matches: 230

Percent Similarity: 75.43% Conservative: 31

Best Local Similarity: 66.47% Mismatches: 65

Query Match: 62.92% Indels: 21

DB: 21 Gaps: 4

US-09-857-613a-28 (1-350) x AA298317 (1-1071)

QY 8 Prothrillesercysllehisllehisrhpheargserglnserproargthrphela 27

DB 61 CCACTCTCTTTCGGCTCAAGTCATCGCTTCTTT-CGGTCTCCATCCTCTCTCTCC 119

QY 28 Argileargval---GlyProargsertrpalaProillearglaseralalaser 46

DB 120 TCACTCTATGACGACACCGCTGGAACGCTGCTGCGGCTGCTGCTGCTGCTGCTGCT 179

QY 47 GluarggluilevalleuglnlnsProlysAspaspLysLysleugln 66


```

QY 67 LysGlyIleAlaGluPheTyrAspGluSerSerGlyLeuTrpGluAsnIleTrpGlyAsp 86
DB 246 AAGAGATAGCGGAGTTCACATGAACTCCGGTTTGGGAAAGAGATTGGGAGAT 305
QY 87 HisMetHisHisGlyPheTyrAspSerAspSerThrValSerLeuSerAsp-----His 104
DB 306 CATATCATCATAGCTTTTATGACCTGATTTCTGTTCACTTCTGATTTCTGCTGAC 365
QY 105 ArgAlaAlaGlnIleArgMetIleGlnGluSerLeuArgPheAlaSerValSer-----Glu 123
DB 366 AAGGAGCTCAGATCCGCTATGATGGAAGACTCTCCGTTCCCGGCTGTACTGATGAA 425
QY 124 GluArgSerIleTrpProLysSerIleValAspValGlyCysGlyIleGlyGlySer 143
DB 426 GAGGAGCAGAAAAAGATTAAGAAAGTAGTGATGGTGGGTGGATTTGAGGAGAGCTCA 485
QY 144 ArgTyrLeuAlaLysLysPheGlyAlaThrSerValGlyIlePheLeuSerProValGln 163
DB 486 AGTATATCTTGCCTTAATTTGAGCTGAATTCGATTTGCTACTCTCAGCCCTGTTGAG 545
QY 164 AlaGlnArgAlaAsnAlaLeuAlaAlaAlaGlnGlyLeuAlaAspLysValSerPheGln 183
DB 546 GCCAAGAGAGCCCATATGATCCGGCGCTCCATTCATCTCTCATAGAGCTTCTCTCCAA 605
QY 184 ValAlaAspAlaLeuGlnGlnProPheSerAspGlyGlnPheAspLeuValTrpSerMet 203
DB 606 GTTGGCGATCCGTTCGATTCAGCATTCGAAAGATGAAATTCATCTAGTGTGCTCATG 665
QY 204 GluSerGlyGluHisMetProAspLysAlaLysPheValGlyGluLeuAlaValAla 223
DB 666 GAGAGTGGTAGCATATGCTGACAGGCAAGTTGTATAAGATTTGATGATGCTGCTGCG 725
QY 224 AlaProGlyAlaIleIleIleIleValThrTrpCysHisArgAspLeuGlyProAspGlu 243
DB 726 GCCTCAGGAGAGTGAATTAATATAGTACATGATGCGCATTAATCTATCTCGCGGGGAG 785
QY 244 GlnSerLeuHisProTrpGluGlnAspLeuLeuLysLysIleCysAspAlaIleTyrTrpLeu 263
DB 786 GAAGCTTTCGACCGCTGGGAGCAAAACATCTTCGACAAATCTGTATAGACGTTCTATCTC 845
QY 264 ProAlaTrpSerSerThrSerAspTrpValLysLeuGlnInsLeuSerLeuGlnAsp 283
DB 846 CCGGCTGGTGGCTCCACCGATGATATGCTTCACTTCTCAATCCATCTCTCCAGGAT 905
QY 284 IleLysSerLysAspTrpSerArgPheValAlaProPheTrpProAlaValIleArgSer 303
DB 906 ATTAAGTGTCCGATTTGCTCAGAGAACGTAGCTCTTCTGGCTGGGTTATATCGGACT 965
QY 304 AlaPheThrTrpLysGlyLeuSerSerLeuLeuSerSerGlyGlnLysThrIleLysGly 323
DB 966 GCATTACATGAAAGGCGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1025
QY 324 AlaLeuAlaMetProLeuMetIleGluGlyTyrLysLysAspLeuIleLysPheAlaIle 343
DB 1026 GCATTGTCAATGCCATTGATGATTGAAGTTACAAAGAGCTGTCAATTAACTTTGGTATC 1085
QY 344 IleThrCysArgLysPro 349
DB 1086 ATCACTTCCCAAGACCA 1103

```

```

XX OS Arabidopsis sp.
XX FH Key Location/Qualifiers
XX CDS 207..1253
XX FT /tag= "a"
XX FT /gene= "165H5T7"
XX FT /product= "gamma-TMT"
XX PN W09904622-A1.
XX PD 04-FEB-1999.
XX XX
XX PF 22-JUL-1998: 98W0-US15137.
XX PR 17-JUL-1998: 98US-0053819.
XX PR 25-JUL-1997: 97US-0053819.
XX PR 26-JAN-1998: 98US-0072497.
XX PA (UYNE-) UNIV NEVADA.
XX PI Dellapenna D, Shintani DK;
XX DR MPI: 1999-142458/12.
XX DR P-PSDB: AAW95017.
XX PT Newly isolated DNA fragment comprising a ^c-tocopherol (vitamin E)
XX PT methyltransferase coding sequence - useful for producing
XX PT ^a-tocopherol, and transgenic plants, seeds and oils with an altered
XX PT tocopherol profile
XX PS Claim 2; Page 36-38; 46pp: English.
XX XX
XX CC The invention provides DNA sequences encoding gamma-tocopherol
XX CC methyltransferases (gamma-TMT). The gamma-TMT encoding genes SLR0089
XX CC and 165H5T7 are isolated from Synechocystis and Arabidopsis species
XX CC respectively. The DNA fragments are useful for producing transgenic
XX CC plants containing a higher alpha-tocopherol:gamma-tocopherol ratio. They
XX CC are also useful for producing alpha-tocopherol and plants with increased
XX CC gamma-tocopherol, which may be useful in certain industries such as the
XX CC meat industry e.g. for developing forage plants to feed animals. The
XX CC production of transgenic plants (and seeds) with a higher alpha-
XX CC tocopherol:gamma-tocopherol ratio, using the gamma-TMTs, will increase
XX CC the level of alpha-tocopherol in the human diet, in addition to enhancing
XX CC the stability and shelf life of plants and plant products. Increased
XX CC levels of alpha-tocopherol will also increase meat quality and extend
XX CC shelf life of post-processed meat products. Plants with a higher alpha-
XX CC tocopherol:gamma-tocopherol ratio may also produce advantageous
XX CC phenotypes. The present sequence represents the 165H5T7 gene encoding the
XX CC Arabidopsis gamma-TMT.
XX SQ Sequence 1790 BP; 474 A; 404 C; 436 G; 476 T; 0 other;
XX XX
XX XX Alignment Scores:
XX Pred. No.: 1,35e-109 Length: 1790
XX Score: 1151.50 Matches: 230
XX Percent Similarity: 75.43% Conservative: 31
XX Best Local Similarity: 66.47% Mismatches: 65
XX Query Match: 62.92% Indels: 21
XX DB: 20 Gaps: 4
XX XX
XX US-09-857-613a-28 (1-350) x AAX17789 (1-1790)
XX QY 8 ProThrIleSerCysIleHisIleHisThrPheArgSerGlnSerProArgTrpPheAla 27
XX DB 259 CCAACTCTCTTTCGGCTCAAGTACATCGCTTCTTT-CGGTCCATGCTCTCTCTCC 317
XX QY 28 ArgIleArgVal---GlyProArgSerTrpAlaProIleArgAlaSerAlaSerSer 46
XX DB 318 TCACGTCTATAGCAGCAACCGCTGGAACCTGCTGGCGCTGCTGTACTATCCACT 377
XX QY 47 GluArgGlyGluIleValLeuGlnGlnLysProLysLysAspAspLysLysLysLeuGln 66
XX DB 111

```

RESULT 8
 AAX17789
 ID AAX17789 standard; DNA: 1790 BP.
 XX
 AC AAX17789;
 XX
 DT 21-MAY-1999 (first entry)
 XX
 DE Arabidopsis gamma-tocopherol methyltransferase encoding gene.
 XX
 KW Gamma-tocopherol methyltransferase; gamma-TMT: SLR0089; 165H5T7: meat;
 transgenic plant; alpha-tocopherol; gamma-tocopherol; forage plant; ds.

DB 378 GAG-----GGCCTAAGA 389
QY 67 LysGlyIleAlaGluPheThrAspGluSerSerGlyLeuTrpGluAsnIleTrpGlyAsp 86
DB 390 AAGAGAAATGACGAGCTTCAACAATGAACTCGGTTGGGAGAAAGATTGGGAGAT 449
QY 87 HisMetHisHisGlyPheThrAspSerSerThrValSerLeuSerAsp-----His 104
DB 450 CATATGATCATGATGCTTTTATGACCTGATTTCTTCTGTTCACTTCTGATTTCTGATG 509
QY 105 ArgAlaAlaGlnIleArgMetIleGlnGluSerLeuArgPheAlaSerValSer---Glu 123
DB 510 AAGGAAGTCACAGATCGCTATGATGAAGATCTCTCCGTTCCCGGCTTACTGATGA 569
QY 124 GluArgSerLeuTrpProLysSerIleValAlaGlyCysGlyIleGlyGlySer 143
DB 570 GAGGAGAGAGAAAAAGATAAGAAAGAGTGGAGTGGGATGGGATGGAGAGCTCA 629
QY 144 ArgTyrlleuAlaLysLysPheGlyValThrSerValGlyIleThrLeuSerProValGln 163
DB 630 ACATATCTTGGCTCTTAATTTGGAGCTGAATGATGCTTACTCTCAGCCTGTTCA 689
QY 164 AlaGlnArgAlaAsnAlaLeuAlaAlaGlnGlyLeuAlaAspLysValSerPheGln 183
DB 690 GCCAAGAGAGCCAAATGATCTCCGCGCTCATCTCTCATAGAGCTTCTCTTCA 749
QY 184 ValAlaAspAlaLeuGlnGlnPhePheSerAspGlyGlnPheAspLeuValTrpSerMet 203
DB 750 GTTGGCGATGCGTTGATGATGAGCATTCGAAGATGGAATTCATCATGCTGCTCATG 809
QY 204 GluSerGlyGluHisMetProAspLysAlaLysPheValGlyLeuAlaArgValAla 223
DB 810 GAGAGTGGTGAAGATATGCTGCAAGAGCCAAAGTTTAAAGAGTTGATGATGCTGCG 869
QY 224 AlaProGlyAlaIleIleIleIleValThrTrpCysHisArgAspLeuGlyProAspGlu 243
DB 870 GCTCCAGAGAGTGAATTAATAGTACATGATGATGATGATGATGATGATGATGATG 929
QY 244 GluSerLeuHisProTrpGluGlnAspLeuLysLysIleCysAspAlaTrpTyrLeu 263
DB 930 GAAGCTTTGGAGCGCGTGGAGCAAAACATCTTGACAAAATCTGTAGACGTTCTATCTC 989
QY 264 ProAlaTrpCysSerThrSerAspTrpValLysLeuLeuGlnSerLeuGlnAsp 283
DB 990 CCGGCTTGGTCCACCCAGATGATGATGATGATGATGATGATGATGATGATGATG 1049
QY 284 IleLysSerGluAspTrpSerArgPheValAlaProPheTrpProAlaValIleArgSer 303
DB 1050 ATTAAGTGTGCGGATTTGGTGAAGAAGCTGCTTCTGCGCTGCGGATATGCGGACT 1109
QY 304 AlaPheThrTrpLysGlyLeuSerSerLeuLeuSerGlyGlnLysThrIleLysGly 323
DB 1110 GCATTAACAGGAAGGCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1169
QY 324 AlaLeuAlaMetProLeuMetIleGlnGlyTyrLysLysAspLeuIleLysPheAlaIle 343
DB 1170 GCATTAACAGGAAGGCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1229
QY 344 IleThrCysArgLysPro 349
DB 1230 ATCATTGCCAGAACCA 1247

KW Protein identification: signal transduction pathway;
KW metabolic pathway: promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX
PD EP1033405-A2.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128334.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139470.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.


```

Db 362 ATGGCCATTGATCTTTGTCGGGNCANTGGAGANTGCTGACACATGCGCAACAAACAGA 421
Qy 214 ysphevalglvleuvalaargvalaalaaproglialaallellellevalthrt 234
Db 422 AGTTGTGAAGGAGCTGGCAGCGCTCGCAGCTCCAGAGCAATCATCATCGTGACCT 481
Qy 234 rrcysarhargaspheuglyproaspgluglnserleuhisprotrpoglinspleu 254
Db 482 GGTGCATAGAGAACCTCGCCCATCGAGAGACTGAAACCTACACAGCTGAATCTTT 541
Qy 254 eulysylleecysasplaatyrtrleuproalaatpccysserthserasptyrval 274
Db 542 TGAATAAGATTGTGATGATTAATTAATCTCCGATTTGGTCTCTCCCTCGATTATGCA 601
Qy 274 yslleuenglinsleuserleuglinspilleyssergluasprrpserargphea 294
Db 602 AGATTGCGGAGCATGTCCTCTGTAGAGATATCAAAAGCCGACAGTGTCAAAACGTGG 661
Qy 294 laiphetrtrproalavalileatrgserialaphethrtrpysgllyeuserleu 314
Db 662 CCCCTCTGCGCTCTGTCATCATCATGACAGCACTGACATGAAAGCCCTCATCTCTAC 721
Qy 314 euserserglinslythrtrleuysglialaleualaameproumetilegllyt 334
Db 722 TAGAGAGTGTGATGAGAGACATTAAGAGGACACTGTGATGCTCTCATGATCAAGCT 781
Qy 334 yrllysaspheulleysphe-alaallellethrCysarglysprou 350
Db 782 ACAAGAAAGCGCTCATTAAGTTCAAGCATCATCACTGCGCAAAACCCCA 832

RESULT 12
AAA29151
ID AAA29151 standard: cDNA: 792 BP.
AC AAA29151:
XX 12-SEP-2000 (first entry)
DE Corn gamma tocopherol methyltransferase partial coding sequence.
XX Vitamin E: alpha-tocopherol; biosynthesis; enzyme; inhibitor; herbicide;
KW gamma-tocopherol methyltransferase; ss.
XX Zea mays.
OS
XX Location/Qualifiers
FH 2..574
FT CDS
FT /tag= a
FT /product= gamma_tocopherol_methyltransferase
FT /partial
XX MO200032757-A2.
XX 08-JUN-2000.
XX 02-DEC-1999: 99WO-US28588.
XX 03-DEC-1998: 98US-0110781.
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX Cahoon RE, Coughlan SJ, Miao G, Rafalski JA:
XX WPI: 2000-412309/35.
XX P-PSDB: AAY96459.
XX Poly nucleotide encoding a vitamin E biosynthetic enzyme polypeptide
XX such as gamma tocopherol methyltransferase, useful for synthesizing
XX vitamin E or as a target for design and discovery of herbicides
XX Claim 3; Page 49; 82pp; English.

```

```

CC AAA29151-69 encode vitamin E (alpha-tocopherol) biosynthetic enzymes or
CC their fragments. The cDNA clones were identified by BLAST searches based
CC on similarity to Synecocystis sp. and Arabidopsis thaliana vitamin E
CC biosynthetic enzymes. The enzymes are useful for synthesizing
CC plastoquinones or tocopherols (especially vitamin E). Vitamin E is
CC required for plant growth, therefore the enzymes may be used for the
CC discovery of new herbicides. The enzymes can be used in methods to
CC evaluate potential inhibitors, which may have use as herbicides.
CC Additionally, transgenic expression of, e.g. gamma-tocopherol
CC methyltransferase affords the ability to manipulate tocopherol levels as
CC desired for a particular application.
XX
SQ Sequence 792 BP: 201 A: 204 C: 215 G: 166 T: 6 other:
Alignment Scores:
Pred. No.: 2.77e-71 Length: 792
Score: 778.00 Matches: 145
Percent Similarity: 85.86% Conservative: 19
Best Local Similarity: 75.92% Mismatches: 27
Query Match: 42.51% Indels: 0
Gaps: 0
DB: 21
US-09-857-613A-28 (1-350) x AAA29151 (1-792)
Qy 159 leuserprovalglinalaglnarvalaasnaalaleualaalaaglinclyleualaasp 178
Db 2 TTGAGCCCTGTTCAAGCCGAGAGAGAAATGCTTCCTCGCCAGCCGAGGGTTGCGCAT 61
Qy 179 lysvalserpheglvalaalaasplaleuglinsprrpserargphea 198
Db 62 CAGGTTACTCTGCAAGTTGCTGATGCTGAGCAACCGTTCTCGAGCGGAGTTGAT 121
Qy 199 leuvaltrpsermetgluserlglunhismetprouasphevalaaleglu 218
Db 122 CTGGCTGTGTCATGAGAGAGTGGCGAGCAGTCCGAGCAAGAAATTTTGTAGTGAG 181
Qy 219 leualaargvalaalaalaproglialaallellellellevalthrtprcysarhargasp 238
Db 182 CTAGCACCGGCTGGCGCTCTCTGAGGACATTAATCATGTGATGCTCATGAGATGCAAC 241
Qy 239 leuglyproaspgluglnserleuhisprotrpoglinspleuulsylsilecys 258
Db 242 CTGATCCATCCGAAACCTCGCTAAAGCCGATGAACCTGAGCTCTGAGAGATATGC 301
Qy 259 aspalaatyrtrleuproalatrprcysserthrserasptyrvalysleuenglins 278
Db 302 GACGCTACTACCTCCGAGACTGTCTCACTTCACTATGTGAACTTCCAGTCA 361
Qy 279 leuserleuglinspilleyssergluasprrpserargpheaalaprophetrpro 298
Db 362 CTGTCTCTGAGGATATCAAGACAGCTGACTGTGCGAGACGTGCCCCGTTTG6CCC 421
Qy 299 alaavallleargseralaphethrtrpysgllyeuserleuuserleuserglgln 318
Db 422 GCCGTATAAATCAACCGCTACATGAGAGGCTTCACTCTCTGTCGACAGCCGATG 481
Qy 319 lysthrilleysglialaleualaametproleumetilegllytrlylsysasphe 338
Db 482 AAGACGATCAGAGCGCGATGATGCGCTAATGATCAGGCGTAAAGAAAGGGGCTC 541
Qy 339 lleysphealallellethrCysarglysprou 349
Db 542 ATCAAAATTCACCATCATCACTGTGCAAGCTT 574

RESULT 13
AAK17788
ID AAK17788 standard: DNA: 954 BP.
XX
XX AAK17788:
AC AAK17788:
XX 21-MAY-1999 (first entry)
XX Synecocystis gamma-tocopherol methyltransferase encoding gene.

```

```

XX  Gamma-tocopherol methyltransferase; gamma-TMT; SLR0089; 165H5T7; meat;
KM  transgenic plant; alpha-tocopherol; gamma-tocopherol; forage plant; ds.
XX
OS  Synecocystis sp.
FH  Key      Location/Qualifiers
FT  CDS      1..954
FT           /*tag= "a"
FT           /gene= "SLR0089"
FT           /product= "gamma-TMT"
XX
XX  W09904622-A1.
XX  04-FEB-1999.
XX
XX  22-JUL-1998; 98WO-0515137.
XX
XX  17-JUL-1998; 98US-0053819.
XX  25-JUL-1997; 97US-0053819.
XX  26-JAN-1998; 98US-0072497.
XX
XX  (UYNE-) UNIV NEVADA.
XX
XX  Dellapenna D, Shintani DK;
XX
XX  WPI; 1999-142458/12.
XX  P-PSDB; AAM95016.
XX
XX  Newly isolated DNA fragment comprising a ^c-tocopherol (vitamin E)
XX  methyltransferase coding sequence - useful for producing
XX  ^a-tocopherol, and transgenic plants, seeds and oils with an altered
XX  tocopherol profile
XX
XX  Claim 2; Page 33-35; 46pp; English.
XX
XX  The invention provides DNA sequences encoding gamma-tocopherol
XX  methyltransferases (gamma-TMT). The gamma-TMT encoding genes SLR0089
XX  and 165H5T7 are isolated from Synecocystis and Arabidopsis species
XX  respectively. The DNA fragments are useful for producing transgenic
XX  plants containing a higher alpha-tocopherol:gamma-tocopherol ratio. They
XX  are also useful for producing alpha-tocopherol and plants with increased
XX  gamma-tocopherol, which may be useful in certain industries such as the
XX  meat industry e.g. for developing forage plants to feed animals. The
XX  production of transgenic plants (and seeds) with a higher alpha-
XX  tocopherol:gamma-tocopherol ratio, using the gamma-TMTs, will increase
XX  the level of alpha-tocopherol in the human diet, in addition to enhancing
XX  the stability and shelf life of plants and plant products. Increased
XX  levels of alpha-tocopherol will also increase meat quality and extend
XX  shelf life of post-processed meat products. Plants with a higher alpha-
XX  tocopherol:gamma-tocopherol ratio may also produce advantageous
XX  phenotypes. The present sequence represents the SLR0089 gene encoding a
XX  Synecocystis gamma-TMT.
XX
XX  Sequence 954 BP; 204 A; 232 C; 271 G; 247 T; 0 other:
XX
XX
XX  Alignment Scores:
XX  Pred. No.: 1.31e-57 Length: 954
XX  Score: 647.50 Matches: 135
XX  Percent Similarity: 63.168 Conservative: 45
XX  Best Local Similarity: 47.378 Mismatches: 102
XX  Query Match: 35.388 Indels: 3
XX  Gaps: 3
XX
US-09-857-613a-28 (1-350) x AAX17788 (1-954)
OY 65 LeuGLuLyGLyLLeaLagLuPheTYrAspGLuSerSerGlyLeuTrpGLuAsnLleTrp 84
DB 100 CTCTACGAAATAATTAATAATTCTACGACGACTCCAGCGCTCTCGGAAGACGTTTG 159
OY 85 GLAspHisMetHisGlyPheTYrAspSerAspSerThraValSerLeuSerAspHis 104
DB 160 GGTGAGCATATGACACCGCTACTACGGTCCACGCGACCTATCGGATC---GATCGC 216

```

```

OY 105 ArgAlaIaGlnLleArgMetLleGlnGlnSerLeuArgPheAlaSerValSerGlnLu 124
DB 217 CGCCAGGCTCAATGATCTGATCAAGAACTATTGGCTGGCA---GTGCCCCAAT 273
OY 125 ArgSerLysTrpProLysSerLleValAspValGlyCysGlyLleGlyLysSerArg 144
DB 274 AGCGCCAA---CCACGAAATAATTCATGATTAGGCTGTGGCATTTGGCGGAGTTTG 330
OY 145 TyrLeuAlaLysLysPheGlyAlaThrSerValGlyLleThrLeuSerProValGlnAla 164
DB 331 TACTGGCCCAACCAACCAACCAAGATGATGGGGCTAGCTTTCCCACTGGCAGTGC 390
OY 165 GlnArgAlaAsnAlaLeuAlaAlaLagGlnGlyLeuAlaAspLysValSerPheGlnVal 184
DB 391 GAACGGCGGGGAAAGGCCCGGCTGGGGTGGCTCAACCTGCCAGTTTGAGGG 450
OY 185 AlaAspAlaLeuGlnGlnProPheSerAspGlyGlnPheAspLeuValTrpSerMetGlu 204
DB 451 GCCAATGCTTGATTTGGCTTGTCTGCCATTCCTTTGACTGGGTTGCTGCGA 510
OY 205 SerGlyGlnHisMetProAspLysAlaLysPheValGlyLleLeuAlaArgValAlaAla 224
DB 511 ACTGGGAGACACATGGCCCAACAAGCTAGTTTACAGAAGCTTGGCGGACTTAA 570
OY 225 ProGlyAlaLleLleLleLleValThrTrpCysHisArgAspLeuGlyProAspGlnGln 244
DB 571 CCAGTGGCCGCTGATTTTACGACCTGTGTCAATGCCATTCATTCATTCATTCATTC 630
OY 245 SerLeuHisProTrpGlnGlnLeuLysLysLleCysAspLalTrpTrpLeuPro 264
DB 631 CCCTGACTGCCGATGACGTCGCCATTCACCAAGCATGTATGACCTTACTGTTGGCC 690
OY 265 AlaTrpCysSerThrSerAspTyrValLysLeuGlnSerLeuSerLeuGlnAspLle 284
DB 691 TATGCTGTTCCCTGCCGACGACGACGACGACGACGACGACGACGACGACGACGAC 750
OY 285 LysSerGlnAspTrpSerArgPheValAlaProPheTrpProAlaValLleArgSerAla 304
DB 751 AATACGCGCATGATGCAATGCGGCGGACCTTTTGGACCGGCTGATGATGCTTGGC 810
OY 305 PheThrTrpLysGlyLeuSerSerLeuSerGlyGlnLysThrLleLysGlyAla 324
DB 811 TTCGATCCCCGGGTGTGTGGGCTTGTGGGCAAGCGGCAAAATTTATCATGCGCC 870
OY 325 LeuAlaMetProLeuMetLleGlnGlyTyrLysLysAspLeuLleLysPheAlaLleLle 344
DB 871 CTGTGTTTACGATTATGAATGGGCTATGAACGGGATTTAGCGTTTGGCTTATTA 930
OY 345 ThrCysArgLysPro 349
DB 931 ACGGGATTAAGCCT 945
XX
XX  RESULT 14
XX  ID AAF26173 standard; DNA; 930 BP.
XX
XX  AAF26173;
XX
XX  26-APR-2001 (first entry)
XX
XX  Synecocystis PCC6803 MPMT encoding DNA SEQ ID 7.
XX
XX  2-methyl-6-phytylhydroquinone methyltransferase; MPMT; transgenic plant;
XX  tocopherol; tocotrienol; vitamin E; 2,3-dimethyl-6-phytyl hydroquinone;
XX  2,3-dimethyl-6-geranylgeranyl hydroquinone; ds.
XX
XX  Synecocystis sp.
XX
XX  Key      Location/Qualifiers
XX  CDS      10..915
XX           /*tag= "a"
XX           /product= "MPMT"
XX

```


XX Claim 2; Page 30-32; 34pp; English.
PS
XX The present sequence encodes a 2-methyl-6-phytylplastoquinol/2-methyl-
CC 6-solaniplastoquinol-9-methyltransferase (methyltransferase 1) protein.
CC The enzyme isolated from *Synechocystis* strain PCC6803. The enzyme is
CC a fundamental enzyme in the production of tocopherols and plastoquinols
CC in higher plants. The enzyme is believed to be involved in regulating
CC the relative amounts of the various tocopherols in photosynthetic
CC organisms. Genetic constructs comprising the methyltransferase 1 coding
CC region under the control of a plant promoter are used to produce
CC transgenic plants, to alter the ratio of delta-tocopherol:gamma-
CC tocopherol in plants. This increases the nutritive value of the plants
CC and products produced from them for humans and animals.
XX

SO Sequence 957 BP; 203 A; 244 C; 268 G; 242 T; 0 other;

Alignment Scores:

Pred. No.:	4,7e-31	Length:	957
Score:	392.00	Matches:	92
Percent Similarity:	51.76%	Conservative:	40
Best Local Similarity:	36.08%	Mismatches:	87
Query Match:	21.42%	Indels:	36
DB:	21	Gaps:	10

05-09-857-613a-28 (1-350) x AA261599 (1-957)

OY 69 IleAlaGluPheTYrAspGlu-----SerSerglyLeuTrpGluAsnIleTrpGlyAsp 86
DB 106 GTGGCCAAAGCGCTTACGACCAATGACAGACAGCGCATTTGGAAATATTACTGGGGCCAC 165
OY 87 HlsmethlshlglyPheTYrAspSerSerpThrValSerLeuSerAspHisArgAla 106
DB 166 CATATCCACCTCGGCATTAT---GGCATCCGCCACGTGCCCAAGCAT----- 210
OY 107 AlaGlnIleArgMetIleGlnGluSerLeuArgPheAlaSerValSergluGluArgSer 126
DB 211 -----TTCATCCAAATCGAAATGATTTGTC-----CATGCCATGGCC 249
OY 127 LysTrp-----ProlysSerIleValAspValGlyCysGly 138
DB 250 CAGTGGGGCGGATTAGATACACTTCCCGCGCACACGCTATTGGATGTGGGTCCGCC 309
OY 139 IleGlyGlySerSerpArgTYrLeuAlaLysLysPheGlyAlaThrSerValGlyIleThr 158
DB 310 ATTGGCGGTAGACGATCTCGCCAAAGATTAAGTTTAACGTTACCGCATCAC 369
OY 159 LeuSerProValGlnAlaGlnArgAlaAsnAlaLeuAlaAlaGlnGlyLeuAlaAsp 178
DB 370 ATTAGTCCCAACAGTGAACAGGGCGGACGGAATTAACCTCCCGCATGTGACGGCC--- 426
OY 179 LysValSerPheGlnValAlaAspAlaLeuGlnGlnProPheSerAspGlyGlnPheAsp 198
DB 427 -----AAGTTTGGCGGTAGACATGCTTGTCTTTCTGTACGCGTAGTTCCGAC 480
OY 199 LeuValTrpSerMetGluSerGlyLysIleHisMetProAspLysAlaLysPheValGlyLeu 218
DB 481 GTAGTTGTGCTGCTGAGACAGGGCCCAATGCTGACAAAGCTGTGTTGCCAAGGAA 540
OY 219 LeuAlaArgValAlaAlaProGlyAlaIleIleIleValAlaThrTrpCysHisArgAsp 238
DB 541 TTACTGCGGTCGTGAACACAGGGGCGCATTTGTGTGGCGGATTCGAAATCAACGGGAC 600
OY 239 LeuGlyProAspGluGln---SerLeuHisProTrpGluGlnAspLeuLeuLysIle 257
DB 601 -----GATGCCCAAGTCCCTCACTTGCGGAAAAACAGATGATCGACACATG 651
OY 258 CysAspAlaTrpTYrLeuProAlaTrpCysSerThrSerAspTYrValLysLeuGln 277
DB 652 TTGGATCAATGCTCCACCTCTTGGCAGCATTTGAAGCTTTGGCGAAATTTGGA 711
OY 278 SerLeuSerLeuGlnAsp-----IleLysSerGluAspTrpSerArgPheValAlaPro 295

DB 712 GCCACGGCTTTGGAGAGGCCAGGTGACTACTGCTGATTGGACTGTACCGACCTCCCC 771
OY 296 PheTrpProAlaValIleArgSerAlaPheThrTrpLysGlyLeu 310
DB 772 GCTTGGTTGATACCATTT-----TGCAGGGCATTT 801

Search completed: March 19, 2003, 07:20:17
Job time : 234 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 19, 2003, 07:11:35 ; Search time 49 Seconds
(without alignments)
2190.548 Million cell updates/sec

Title: US-09-857-613A-28
Perfect score: 1830
Sequence: 1 MATVVRIPITISCIHIIFERS.....IEGYKDLIKFAITCRKPE 350

Scoring table:		
BLOSUM62		
Xgapop	10.0	Xgapext 0.5
Ygapop	10.0	Ygapext 0.5
Fgapop	6.0	Fgapext 7.0
Delop	6.0	Delext 7.0

```
Searched:      441362 seqs, 153338381 residues
Total number of hits satisfying chosen parameters:  882724
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```

Command line parameters:
-MODEL=frim+ p2n. model -DEV=xlh
-o-/sgn2.1/USPO.spool/US09857613/running.12032003-100037-6041/app_query.fasta.1.519
-DB=Issued_Patents.NA -QMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOEXT=0 -UNITS=bits -STAR=1 -END=1 -MATRIX=blosome2 -TRANS=Illum40.cdi
-LIST=45 -DOCALLIGN=20 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPFX=prc -NORM=ext -HEAPSIZE=2000000000 -MINLEN=6 -MAXLEN=2000000000
-USER=US09857613 -ECGN=1.1 32 -urnat.12032003.100037.6041 -NCPX=6 -ICPU=3
-NO_XLPRX -NO_MAP -LARGEXURY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
WARN_TIMEOUT=30 -THRADES=1 -XGAP=10 -XGAPEXT=0.5 -FEAP=6 -FEAPEXT=7
-XGAP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

```
Database :
Issued_Patents_MN:*
1: /cgn2_6/prodata/1/lna/5A_COMB.seq:*
2: /cgn2_6/prodata/1/lna/5B_COMB.seq:*
3: /cgn2_6/prodata/1/lna/6A_COMB.seq:*
4: /cgn2_6/prodata/1/lna/6B_COMB.seq:*
5: /cgn2_6/prodata/1/lna/PTUS_COMB.seq:*
6: /cgn2_6/prodata/1/lna/backfiles1.seq:
```

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	392	21.4	957	4	US-09-382-906A-1	Sequence 1, Appl1
2	261	14.3	2122	4	US-09-029-603-1	Sequence 1, Appl1
3	225	12.3	6085	3	US-09-029-603-4	Sequence 4, Appl1
4	220.5	12.0	80161	3	US-09-036-987A-1	Sequence 1, Appl1
5	220.5	12.0	80161	4	US-09-370-700-1	Sequence 1, Appl1
6	189	10.3	1400	4	US-09-041-718-1	Sequence 1, Appl1
7	185	10.1	45377	4	US-08-764-233A-1	Sequence 1, Appl1
8	156.5	8.6	1149	2	US-08-844-305-1	Sequence 1, Appl1
9	152.5	8.3	80161	3	US-09-036-987A-1	Sequence 1, Appl1
c 10	152.5	8.3	80161	4	US-09-370-700-1	Sequence 1, Appl1
c 11	138	7.5	4833	4	US-09-066-047-1	Sequence 1, Appl1
12	137.5	7.5	3236	4	US-08-961-537-222	Sequence 222, Appl1

13	131	7.2	8051	2	US-08-576-626A-2	Sequence 2, Appl
14	127	6.9	403765	4	US-09-103-840A-2	Sequence 2, Appl
15	120	6.6	3719	1	US-08-920-882-10	Sequence 10, Appl
16	120	6.6	3719	1	US-08-920-882-10	Sequence 10, Appl
17	120	6.6	3719	1	US-08-921-177-10	Sequence 10, Appl
18	120	6.6	3719	1	US-08-362-377C-10	Sequence 10, Appl
19	120	6.6	3719	2	US-08-920-828-10	Sequence 10, Appl
20	117.5	6.4	4411529	4	US-09-103-840A-1	Sequence 11, Appl
21	114.5	6.3	2451	3	US-09-217-609A-21	Sequence 21, Appl
22	114.5	6.3	2451	4	US-08-873-235B-21	Sequence 21, Appl
23	113.5	6.3	4403765	4	US-09-103-840A-2	Sequence 2, Appl
24	113.5	6.2	741	3	US-09-217-609A-29	Sequence 29, Appl
25	113.5	6.2	741	4	US-08-873-235B-29	Sequence 29, Appl
26	109	6.0	9377	4	US-09-221-017B-11002	Sequence 1002, App
27	108.5	5.9	2077	4	US-09-221-017B-821	Sequence 821, App
28	108.5	5.9	4463	2	US-08-760-489-1	Sequence 1, Appl
29	108.5	5.9	4463	2	US-08-760-489-3	Sequence 3, Appl
30	108.5	5.9	4463	4	US-09-185-373-1	Sequence 1, Appl
31	108.5	5.9	4463	4	US-09-185-373-3	Sequence 3, Appl
32	107.5	5.9	8257	4	US-09-484-970B-65	Sequence 65, Appl
33	104	5.7	756	2	US-08-401-068-11	Sequence 11, Appl
34	104	5.7	756	2	US-08-846-338-11	Sequence 11, Appl
35	104	5.7	5872	3	US-08-411-768B-1	Sequence 1, Appl
36	104	5.7	5872	3	US-08-411-768B-6	Sequence 6, Appl
37	103	5.6	4911	4	US-09-718-852-1	Sequence 1, Appl
38	103	5.6	4911	4	US-09-718-852-1	Sequence 1, Appl
39	103	5.6	4911	4	US-09-718-852-1	Sequence 1, Appl
40	102.5	5.6	5100	1	US-08-457-245-1	Sequence 1, Appl
41	102.5	5.6	4411529	4	US-09-103-840A-1	Sequence 1, Appl
42	101.5	5.5	5822	3	US-08-899-595-4	Sequence 4, Appl
43	101.5	5.5	5822	3	US-08-899-595-5	Sequence 5, Appl
44	101	5.5	1406	4	US-08-936-165A-119	Sequence 119, App
45	99.5	5.4	777	4	US-09-134-001C-407	Sequence 407, App

ALIGNMENTS

```

1      RESULT 1
2      US-09-382-906A-1
3      : Sequence 1, Application US/09382906A
4      : Patent No. 6448475
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT: Dellapenna, Dean
9      : APPLICANT: Shintani, David
10     : TITLE OF INVENTION: Manipulation of Tocopherol Levels in Transgenic Plants
11     :
12     : FILE REFERENCE: 920905,90032
13     :
14     : CURRENT APPLICATION NUMBER: US/09/382,906A
15     :
16     : CURRENT FILING DATE: 1999-08-25
17     :
18     : PRIOR APPLICATION NUMBER: 60/097,863
19     :
20     : PRIOR FILING DATE: 1998-08-25
21     :
22     : NUMBER OF SEQ. ID NOS.: 4
23     :
24     : SOFTWARE: PatentIn Ver. 2.1
25     :
26     : SEQ. ID NO. 1
27     :
28     : LENGTH: 957
29     :
30     : TYPE: DNA
31     :
32     : ORGANISM: Synechocystis PCC6803
33     :
34     : FEATURE:
35     :
36     : NAME/KEY: CDS
37     :
38     : LOCATION: (1)..(954)
39     :
40     : US-09-382-906A-1

```



```

: Sequence 4, Application US/09029603
: Patent No. 6210935
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Engel, Natalie
: APPLICANT: Bietenhader, Jurg
: APPLICANT: Toupet, Christine
: APPLICANT: Pospiech, Andreas
: TITLE OF INVENTION: Staurosporin Biosynthesis Gene Clusters
: FILE REFERENCE: 4-20555/4/PCT
: CURRENT APPLICATION NUMBER: US/09/029, 603
: CURRENT FILING DATE: 1998-03-20
: EARLIER APPLICATION NUMBER: PCT/EP96/03643
: EARLIER FILING DATE: 1996-08-19
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4
: LENGTH: 6085
: TYPE: DNA
: ORGANISM: Streptomyces longisporoflavus
: FEATURE:
: NAME/KEY: misc_RNA
: LOCATION: (378)..(1665)
: OTHER INFORMATION: ORF
: FEATURE:
: NAME/KEY: misc_RNA
: LOCATION: (1174)..(2553)
: OTHER INFORMATION: ORF
: FEATURE:
: NAME/KEY: misc_RNA
: LOCATION: (2593)..(4011)
: OTHER INFORMATION: ORF
: FEATURE:
: NAME/KEY: misc_RNA
: LOCATION: (4013)..(4999)
: OTHER INFORMATION: ORF
: FEATURE:
: NAME/KEY: misc_RNA
: LOCATION: (5071)..(6085)
: OTHER INFORMATION: ORF
: US-09-029-603-4

```

```

Alignment Scores:
Pred. No.: 7.93e-16 Length: 6085
Score: 225.00 Matches: 75
Percent Similarity: 41.45% Conservative: 51
Best Local Similarity: 24.67% Mismatches: 120
Query Match: 12.30% Indels: 58
Gaps: 12
US-09-857-613A-28 (1-350) x US-09-029-603-4 (1-6085)

```

```

QY 33 ProArgSerTrpAlaProIleArgAlaSerAlaIleSer-----GluArgGlyGlu 50
DB 1684 CCCTGCGCCCTGCGACCCCTCGACCCCGGCGCTCGACTCGGACGACGAGAGGACACA 1743
QY 51 IleValLeuGluGlnIleProIleProIleProIleProIleProIleProIleProIle 70
DB 1744 TCCATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1785
QY 71 GluPheTrpAspGlu-----SerSerGlyLeuTrpGluAsnIleTrpGlyAspHisMet 88
DB 1786 GAGGTGTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1842
QY 89 HisHisGlyPheTrpAspSerAspSerTrpValSerLeuSerAspHisArgAlaAlaGln 108
DB 1843 CACGTGGGTACTGGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1902
QY 109 IleArgMetIleGlnIleSerLeuArgPheAlaSerValSerIleGluArgSerIleTrp 128
DB 1903 ACCGATCTGTGTGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1938
QY 129 ProIleSerIleValAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 148

```

```

DB 1939 GACCGGATCTGTGTGAGCTGGGCTGCGCATCGGCGCTGCGGCTCTAGGATTCGCGGA 1998
QY 149 LysPheGlyAlaThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaAsn 168
DB 1999 GCGCAGACGCTCGGCTCACCGGATCACCGGATCACCGGATCACCGGATCACCGGATCACCGGATCACCGGAT 2058
QY 169 AlaLeuAlaAlaGlnGlyLeuAlaAspValSerPheGlnValAlaAspAlaLeu 188
DB 2059 GAGCGGGCGGTGAGTCCGATCGATCGGCGGCGGTCTCTCCCTTCGCGGCTGCGGACCGCATG 2118
QY 189 GlnGlnProPheSerAspGlyGlnPheAspLeuValIleTrpSerMetGluSerIleGluHis 208
DB 2119 GACCTCCCTTCGAGGACGCTCTCTTCGAGGCGGCGCTTCGATCGATCGATCGATCGATCGATCGATCGAT 2178
QY 209 MetProAspValAlaLysPheValGlyGluLeuAlaArgValAlaAlaProGlyAlaIle 228
DB 2179 CTGCGCGACACACACACCGCGCTCAAGGAGATCACCGGCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2238
QY 229 IleIleIleValIleThrTrpCysHisArgAsp-----LeuGlyProAspGluGlnSerLeu 246
DB 2239 CTGCTCATCGCGACCTGTGTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 2298
QY 247 His-----ProIleGluGlnAspLeu 253
DB 2299 GACGGATGCTCTGTATGACGATCGCGGAGATCACACACACCTACGACGAT----- 2352
QY 254 LeuLysLysIleCysAspAlaIleTrpIleProAlaTrpCysSerThrSerAspIleVal 273
DB 2353 CGCGCGCGACTGGCGGAGCG-----GGCTGC----- 2379
QY 274 LysLeuLeuGlnSerLeuSerLeuGlnAspIleLysSerGluAspTrpSerArgPheVal 293
DB 2380 -----GAAGCTGTGAGCTGACGACGATCGGACGACGATCGGACGACGATCGGACGACGATCGGACGACGAT 2415
QY 294 AlaProPheTrpProAlaValIleArgSerAlaPheThrTrpLysGlyLeuSerSerLeu 313
DB 2416 CGGCGCTACTACGCG-----CATGCCGCGCGCGCGCTGCGGCGGCTGCGCGGCGCT 2466
QY 314 LeuSerSerGly 317
DB 2467 CTCGACGCGCGC 2478

RESULT 4
US-09-036-987A-1
: Sequence 1, Application US/09036987A
: Patent No. 6143526
: GENERAL INFORMATION:
: APPLICANT: Baltz, Richard H.
: APPLICANT: Crawford, Mary C.
: APPLICANT: Broughton, Kathryn P.
: APPLICANT: Madduri, Krishnamurthy
: APPLICANT: Merlo, Donald J.
: APPLICANT: Treadway, Paul J.
: APPLICANT: Turner, Jan R.
: APPLICANT: Waldron, Clive
: TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
: NUMBER OF SEQUENCES: 39
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dow Agrosciences LLC Patent Department
: STREET: 9330 Zionsville Road
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: USA
: ZIP: 46268
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/036,987A

```


FILING DATE: 09-MAR-1998
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Stuart, Donald R
 REGISTRATION NUMBER: 28,479
 REFERENCE/DOCKET NUMBER: 50,608
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (317)337-4816
 TELEFAX: (317)337-4847
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 80161 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-09-036-987A-1

Alignment Scores:
 Pred. No.: 1.73e-13 Length: 80161
 Score: 220.50 Matches: 78
 Percent Similarity: 43.99% Conservative: 50
 Best Local Similarity: 26.80% Mismatches: 125
 Query Match: 12.05% Indels: 39
 Gaps: 9

US-09-857-613a-28 (1-350) x US-09-036-987A-1 (1-80161)

QY 69 ILEAGLUPHETRYRASPGLUSERSERGLYLEUTRPGLUASNIETRPGLYASP----- 86
 DB 20201 GTTGGCGGAGATGATGACCTGTCACCGCTTGCAGACTCGGCGGCGCCCTGCG 20260
 QY 87 HSMETHISHISGLYPHETRYRASPSPERTHRVALSERLEUSERSPHISARGALA 106
 DB 20261 GCCATCCACCGGCTACTGCGAGACGCGCGCGCTCTGCGACAGCGC-CGCCGA 20319
 QY 107 ALAGLNIETARGMETILEGLIUSERTLEUARGPHEALASERVALSERGLUGLUARGSER 126
 DB 20320 CCGGCTACCGACCTTGTGCGGACG-----GACCGTGCT 20335
 QY 127 LYSTPRPOLYSER-ILEVALASPVALGLYCGLYILEGLYISERSEARATRYTLE 146
 DB 20356 CGATGCGCGCGCTGCGACGCTGCGATGCGGCGTGCACGACACACGCGCTCGCT 20415
 QY 146 WALALYSLSYPHEGLYALATHRSERVALGLYLETTRLEUSERPROVALGLIALGLINAR 166
 DB 20416 CGCGCGCGACACCGCATCCAGATCACCAGCGCTACCGCTCAAGTGGCCAT 20475
 QY 166 GALASNALEUALAALALAGLNGLYLEUALASPLYVALSERPHEGLINVALALAS 186
 DB 20476 CGCCCTGATTCGCGACCGGACGCGGACTAGCCACCGCGGTCTCTCGTGCCTCGA 20535
 QY 186 PALALEUNGINGINPROPHESERASPGLYGINPHEASPLEUVALTRPSEMETGLUSERI 206
 DB 20536 TGGCATTCCTCGCGCTGACCGGACAAATGCTTCACGCGCGCTGCGCATCGATCGCT 20595
 QY 206 YGLUHISMETPROASPLYALATYSRPHAVAGLYLLEUALAARGVALALALAPROGL 226
 DB 20596 GTTGAAGATGTCGGAACCGGACGCTGCGCATCCGGGAAATCTTCAAGTCAACCCGG 20655
 QY 226 YALALELLELLELLEVALTHRTTPCYSHISARGASPLEUNGILYPROASP---GLUGINSE 245
 DB 20656 TGGCATTCCTCGCGCTGACCGGACGCGGCTCAACAGGAGAGCGGCGGATCCGGTGTG 20715
 QY 245 RLEUHSIPROTPGLUGINASPLEUENLYSLYILECYASAPALATYTRYTLEUENPROAL 265
 DB 20716 CGGGACAGGTGCGCGCGCTT-----CGGATCTCGCTGCT----- 20755
 QY 265 ATTPCYSSERTHRSERASPTYRVALLYSLEULEUNGINSERLEUSERLEUGINASPILLEY 285
 DB 20756 -----GAGCAACTCTGGAATCGCTGCGGACGCGGCGGTTCGA 20793
 QY 285 SSETGLUSPTTP-----SERARGPHEVALALAPROPHETPROAL 299

DB 20794 GATCCTCATTTGGAGGACGCTGCTGCGAGACCGGTAATTCAATCGCGGACGCGCA 20853
 QY 299 AVALLIETARGSERALAPETHRTTPGLYLSERISERLEUSERSEARGLYINLY 319
 DB 20854 AGAGCTC-----GCTGGCACACACCGGATCCGACAGTACGCG-----CC 20898
 QY 319 STHRIELYSGLYALALEUALAMETPROLEUMETILEGLYLYTRYLYLSASPLEUL 339
 DB 20899 GGCTGTGCGCGCGCTGCGCGCGCGCTGCGCATTTATGAAATATGCCCCACGACATG-- 20956
 QY 339 ELYSPHEALALELLELLETHRCYSARGLYSPRO 349
 DB 20957 -GGCTATCGATTCGACGCGCGGACGCGC 20986

RESULT 5
 US-09-370-700-1
 : Sequence 1, Application US/09370700
 : Patent No. 6274350
 : GENERAL INFORMATION:
 : APPLICANT: Baltz, Richard H
 : APPLICANT: Broughton, Mary C
 : APPLICANT: Crawford, Kathryn P
 : APPLICANT: Madduri, Krishnamurthy
 : APPLICANT: Treadway, Patil J
 : APPLICANT: Turner, Jan R
 : TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
 : FILE REFERENCE: 50489 DIVI
 : CURRENT FILING DATE: US/09/370,700
 : EARLIER FILING DATE: 1999-08-09
 : EARLIER APPLICATION NUMBER: US 09/36987
 : NUMBER OF SEQ ID NOS: 39
 : SOFTWARE: Patent Ver. 2.0
 : SEQ ID NO 1
 : LENGTH: 80161
 : TYPE: DNA
 : ORGANISM: Saccharopolyspora spinosa
 US-09-370-700-1

Alignment Scores:
 Pred. No.: 1.73e-13 Length: 80161
 Score: 220.50 Matches: 78
 Percent Similarity: 43.99% Conservative: 50
 Best Local Similarity: 26.80% Mismatches: 125
 Query Match: 12.05% Indels: 39
 Gaps: 9

US-09-857-613a-28 (1-350) x US-09-370-700-1 (1-80161)

QY 69 ILEAGLUPHETRYRASPGLUSERSERGLYLEUTRPGLUASNIETRPGLYASP----- 86
 DB 20201 GTTGGCGGAGATGATGACCTGTCACCGCTTGCAGACTCGGCGGCGCCCTGCG 20260
 QY 87 HSMETHISHISGLYPHETRYRASPSPERTHRVALSERLEUSERSPHISARGALA 106
 DB 20261 GCCATCCACCGGCTACTGCGAGACGCGCGCGCTCTCGGACAGCGC-CGCCGA 20319
 QY 107 ALAGLNIETARGMETILEGLIUSERTLEUARGPHEALASERVALSERGLUGLUARGSER 126
 DB 20320 CCGGCTACCGACCTTGTGCGGACG-----GACCGTGCT 20335
 QY 127 LYSTPRPOLYSER-ILEVALASPVALGLYCGLYILEGLYISERSEARATRYTLE 146
 DB 20356 CGATGCGCGCGCTGCGACGCTGCGATGCGGCGTGCACGACACACGCGCTCGCT 20415
 QY 146 WALALYSLSYPHEGLYALATHRSERVALGLYLETTRLEUSERPROVALGLIALGLINAR 166
 DB 20416 CGCGCGCGACACCGCATCCAGATCACCAGCGCTACCGTACGACGAGTGGCCAT 20475
 QY 166 GALASNALEUALAALALAGLNGLYLEUALASPLYVALSERPHEGLINVALALAS 186


```

Db 47058 GAGGCCGCTCAAGTGTCCGATCGCTTTTCACAAAC---TGGGGG-----ATGTTTCAGAG 47108
OY 92 PheTYrAspSerAspSerThrValSerLeuSerAspHisArgAlaIleGlnIleArgMet 111
Db 47109 GAGGTTTATCGTGAATGTTCTTCAACAGCTGCCAGATTACAGACAGACAGACAGCGT 47168
OY 112 IleGlnIle-----SerLeuArgPheAlaSerValSerGlnIleArgSer 126
Db 47169 TATTGGACACACTCTACGCTACACAGCTGAGACAGATTCCGGCCGACAGAAAGTCCG--- 47225
OY 127 LysTPProLysSerIleValAspValGlyCysGlyIleGlyCysSerArgTYrLeu 146
Db 47226 -----CCACGCAAGATCTCGAGCTCGCGTCAACACTGCGAAGGCGCTCAACTCTCT 47279
OY 147 AlaLysLysPheGly---AlaThrSerValGlyIleThrLeuSerProValGlnIleGln 165
Db 47280 TCCCGGATCGAGGCTGCGACACCTTCTGGGCTGCGACTGCTGCGACGACAGCGGTGGAC 47339
OY 166 ArgAlaAsnAlaLeuAlaIleAlaGlnGlyLeuAlaAspLysValSerPheGlnValAla 185
Db 47340 ATCGCCACAGCCCGGCTTCGCGACGCGG---TCGTTGACCTACGTTCAAGG 47390
OY 186 AspAlaLeuGlnGlnProPheSerAspGlyIlePheAspLeuValTrpSerMetGluSer 205
Db 47391 GACGCGGAGACACCTTCTTCGCGAGCGGGAATTCGACCTGCTCATTCATGTCGAGAGC 47450
OY 206 GlyGlnHisMetProAspLysAlaLysPheValGlyGlnLeuAlaArgValAlaIlePro 225
Db 47451 TCGCACACCTACCTTAATGTCGCGCAATCTTCTGAACTGGCGCGGCTGCGACCG 47510
OY 226 GlyAlaIleIleIleIleValThrTrpCysHisArgAspLeuGlyProAspGlnIleSer 245
Db 47511 GCGGCGGCTC-----TTTTCGATGTGGATGTCTTCGCGACAAATCGC--- 47552
OY 246 LeuHisProTrpGlnIleAspLeuLeuLysLysIleCysAspAlaTYrTYrLeuProAla 265
Db 47553 -----TATTCGGTGATGCAG 47567
OY 266 TrpCys-----SerThrSerAspTYrValLysLeuGlnSerLeuGlnAsp 283
Db 47568 AATTGACAGCAGACAGACAGCGGCGAGCTGCTCAAGAGACCGACATCTCGGAG 47627
OY 284 IleLysSerGlnAspTrpSerArgPheValAlaIleProPheTrpProAlaValIleArgSer 303
Db 47628 TACGTAAGAAGAGCTATCCGCGACAGCTGCGCCCGCC----- 47666
OY 304 AlaPheThrTrpLysGlyLeuSerSerLeuLeuSerSerGlyGlnLysThrIleLysGly 323
Db 47667 -----AGCAAGCAAGCAAGCAAGCTGCGAGC 47693
OY 324 AlaLeuAlaMetProLeu-----MetIleGlnGlyTYrTYrLysLys 336
Db 47694 GCCCTCCCGTACCCCGCTCGGAGAGCTTTTTCAGCTCGGTATGATGCGCGCTACGATCC 47753
OY 337 AspLeuIleLysPheAlaIle 343
Db 47754 GAG-----TTTGGCGCTC 47765

```

```

; ORGANISM: Escherichia coli
US-08-844-305-1
Alignment Scores:
Pred. No.: 7,18e-09 Length: 1149
Score: 156.50 Matches: 69
Percent Similarity: 41.45% Conservative: 45
Best Local Similarity: 25.09% Mismatches: 105
Query Match: 8.55% Indels: 56
DB: 2 Gaps: 14

US-09-857-613a-28 (1-350) x US-08-844-305-1 (1-1149)
OY 71 GluPheTYrAspGluSerSerGlyLeuTrpGluAsnIleTrpGlyAspHisMetHis 90
Db 352 GACCATTAAGATTGGGTAAATGACTTGTTCAGCCGATGTTATCCCTCATGACATAT 411
OY 91 Gly-----PheTYrAspSerAspSerThrValSerLeuSerAspHisArgAlaIle 107
Db 412 TCTCGGCTTACTAGGAAGATGCCGAT-----AATCTGGAATCTGCCACAGCGC 462
OY 108 GlnIleArgMetIleGlnGlnIleValSerLeuArgPheAlaSerValSerGlnIleArgSerLys 127
Db 463 AACCTCAAAATGATTTGTGAAAATTGCACTTA-----AAA 498
OY 128 TrpProLysSerIleValAspValGlyCysGlyIleGlyCysSerArgTYrLeuAla 147
Db 499 CCAAGGATGCGGCTACTGATATGCTGCTGGCGGCGGATGCGACATCATGATGCA 558
OY 148 LysLysPheGlyAlaThrSerValGlyIleThrLeuSerProValGlnIleArgValAla 167
Db 559 TCTAATATATGACCTAAGCCGCTGGGCGGCTCACCATTCTCCGACACGCAAAAATGCT 618
OY 168 AsnAlaLeuAlaIleAlaGlnGlyLeuAlaAspLysValSerPheGlnValAlaAspAla 187
Db 619 CAG-----GAACGCTGTGAAGCCTGATGATCACTT----- 651
OY 188 LeuGlnGlnProPheSerAsp-----GlyGlnPheAspLeuValTrpSerMetGluSer 205
Db 652 TTGCTGCAACATTAATGCTGACCTGACGACGACGAGTTGATGCTATTTCTGTGGGATG 711
OY 206 GlyGlnHisMetProAspLys-----AlaLysPheValGlyGlnLeuAlaArgValAla 223
Db 712 TTGAGCAGCCTCGACCGAAGAAATTAACATATCTATTTCCGCTGGTGATGCTAATTTG 771
OY 224 AlaProGlyAlaIleIleIleIleValThrTrpCysHisArgAspLeuGlyProAspGlu 243
Db 772 AAACCGGAGGACATATTCCTGCTCCATACT-----ATCGGTTGCAAAAAA 816
OY 244 GlnSerLeuHis-----ProTrpGlnAspLeuLeuLysLysIleCysAspAlaTYr 261
Db 817 ACCGATCTGAATGTGATCCCTG-----ATTAAATAATAT 852
OY 262 TYrLeuProAlaTrpCysSerThrSerAspTYrValLysLeuGlnSerLeuSerLeu 281
Db 853 ATTTTTCGAAGGCTGCTGCGCCTCT-----CTACGACAGATGCTCATCCACGCAA 906
OY 282 GlnAspIleLysSerGlnAspTrpSerArgPheValAlaIleProPheTrpProAlaValIle 301
Db 907 CCCACCTTTGTGATGGAAGACTGCAATACCTTCGCTCGTCAATAC----- 951
OY 302 ArgSerAlaPheThrTrpLysGlyLeuSerSerLeuLeuSerSerGlyGlnLys----- 319
Db 952 ---GATACATACGTTGATGCGCTGATGACAGATTCCTCGCCGATGCGCAAAATATGCG 1008
OY 320 ---ThrIleLysGlyAlaLeu-----AlaMetProLeuMet 330
Db 1009 GATAACTATAGTGAACGCTTAACGAATGTTTACATATTATC 1051

```

```

RESULT 9
US-09-036-987A-1/c
; Sequence 1, Application US/09036987A
; Patent No. 6143526
; TYPE: DNA

```

GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patti J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
TITLE OF INVENTION: Production
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R.
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 80161 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1
Alignment Scores:
Pred. No.: 2e-05 Length: 80161
Score: 152.50 Matches: 48
Percent Similarity: 38.38% Conservative: 28
Best Local Similarity: 24.24% Mismatches: 77
Query Match: 8.33% Indels: 45
Gaps: 4
US-09-857-613a-28 (1-350) x US-09-036-987A-1 (1-80161)
OY 85 GLYASPHISMETHISHISGLYPHEPTYRASP----- 94
DB 13517 GGGCGTCCCTGACACGAGTTACTGGCGCGGATCGGAGATCCCGTCCACA 13458
OY 95 -----SeraspserThrValserLeuserSerAspHisArgAlaAlaGlnIleArgMetIle 112
DB 13457 CCGTGGTGGATGCTGGCGACCACTGACGAC-----CTGTTCATCCAC 13413
OY 113 GlnGluSerLeuArgPheAlaSerValSerGluGluArgSerLysTrpProLysSerIle 132
DB 13412 AAGGCCCGCTCCCTCCCGGAGCG-----CACCTG 13383
OY 133 ValaspValGlyCysGlyIleGlyCysSerArgTyrLeuAlaLysLysPheGlyAla 152
DB 13382 TTCGACCTGGCTCGGCAATGGCAGCCGCTAGTCCGTGGCGGACGCGCGCTT 13323
OY 153 ThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaAsnAlaLeuAla 172
DB 13322 CGAGTCAACCGAATCACCCTGAACGCCAGCATCTCGCGCGGACCAAGGCTCGCCAAC 13283

OY 173 AlaGlnGlyLeuAlaAspLysValSerPheGlnValAlaAspAlaLeuGlnGlnProPhe 192
DB 13262 GAGACCGGACTGCGCGACGACTTGTAGTGTGATCTGACAGCGGCCCTGCTCCCTAC 13203
OY 193 SeraspGlnPheAspLeuValTrpSerMetGluSerGlyGlnHisMetProAspLys 212
DB 13202 CCGGACGCTTTCTTTCAGCGCGCGCATGGCGATGCTGCTGTCGATGCTGACACG 13143
OY 213 AlaLysPheValGlyGluLeuAlaArgValAlaAlaProGlyAlaIleIleIleVal 232
DB 13142 GCCCGCGGATCGCGGAGCTGACCGAATCTGGAACCGCGCGGCTGCTGCTGGA 13083
OY 233 ThrTrpCysHisArgAspLeuGlyProAspGluGlnSerLeuHisProTrpGluGlnAsp 252
DB 13082 GACATCATCACTCGGTTGCGATCCCGAAGAC----- 13050
OY 253 LeuLeuLysLysIleCysAspAlaTyrTyrLeuProAlaTrpCysSerThrSer 270
DB 13049 -----TACGCGCGGCTTGGACGGCGACGACC 13023
RESULT 10
US-09-370-700-1/c
Sequence 1, Application US/09370700
Patent No. 6274350
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Treadway, Patti J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
FILE REFERENCE: 50489 DIV1
CURRENT APPLICATION NUMBER: US/09/370,700
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/36987
EARLIER FILING DATE: 1998-03-09
NUMBER OF SEQ. ID NOS: 39
SOFTWARE: Patentin Ver. 2.0
SEQ. ID NO. 1
LENGTH: 80161
TYPE: DNA
ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1
Alignment Scores:
Pred. No.: 2e-05 Length: 80161
Score: 152.50 Matches: 48
Percent Similarity: 38.38% Conservative: 28
Best Local Similarity: 24.24% Mismatches: 77
Query Match: 8.33% Indels: 45
Gaps: 4
US-09-857-613a-28 (1-350) x US-09-370-700-1 (1-80161)
OY 85 GLYASPHISMETHISHISGLYPHEPTYRASP----- 94
DB 13517 GGGCGTCCCTGACACGAGTTACTGGCGCGGATCGGAGATCCCGTCCACA 13458
OY 95 -----SeraspserThrValserLeuserSerAspHisArgAlaAlaGlnIleArgMetIle 112
DB 13457 CCGTGGTGGATGCTGGCGACCACTGACGAC-----CTGTTCATCCAC 13413
OY 113 GlnGluSerLeuArgPheAlaSerValSerGluGluArgSerLysTrpProLysSerIle 132
DB 13412 AAGGCCCGCTCCCTCCCGGAGCG-----CACCTG 13383
OY 133 ValaspValGlyCysGlyIleGlyCysSerArgTyrLeuAlaLysLysPheGlyAla 152
DB 13382 TTCGACCTGGCTCGGCAATGGCAGCCGCTAGTCCGTGGCGGACGCGCGCTT 13323


```

: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/961.527
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Brookes, A. Anders
: REGISTRATION NUMBER: 36,373
: REFERENCE/DOCKET NUMBER: PB340P1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 222:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3236 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: US-08-961-527-222

Alignment Scores:
Pred. No.: 6,82e-06 Length: 3236
Score: 137.50 Matches: 59
Percent Similarity: 40.98% Conserved: 41
Best Local Similarity: 24.18% Mismatches: 85
Query Match: 7.51% Indels: 60
DB: 4 Gaps: 10

US-09-857-613a-28 (1-350) x US-08-961-527-222 (1-3236)
QY 42 SerAlaIaSerSerGluArgGlyGlu-----IleValLeuGluGlnIlySpro 57
DB 2260 TCTAGTTATCAATATATAGATTAGATTGTCTAGATAATATCTATTGTGTTATTAATGAA 2319
QY 58 LysLysAspAspLysLysLysLeuGlnLysGlyIleAlaGluPheTyrAspGluSer 77
DB 2320 GAAAAAAGAGAGGTCTCAAAAGTCCAGAGCAGTCATAGTTTACGAAAT----- 2373
QY 78 GlyLeuTPrGluAsnIleTPrGlyAspHisMetHisGlyPheTyrAspSer 97
DB 2374 -----TCGGGAAAAAAGCGTTAC----- 2391
QY 98 ThrValSerLeuSerAspHisArgAlaIleGlnIleArgMetIleGlnIleuSerLeuArg 117
DB 2392 -----GTCAGCTGGAAGCGTCCACAGATTGG-TTAAFTGCAGAA----- 2432
QY 118 PheAlaSerValSerGluGluArgSerLysTPrProLysSerIleValAspValGlyCys 137
DB 2433 ---GGAGGATTTTCAAAAGAA-----AAGGATACTAGAGGTTGGGTGT 2474
QY 138 GlyIleGlyLysSerArgTyrTleuAlaLysLysPheGlyAlaThrSerValGlyIle 157
DB 2475 AATAGGGGAACACCAATTCAGTTGGCACAGCGTTTGTGCAAGATAACTGCGTGT 2534
QY 158 ThrLeuSerProValGlnIleGlnArgAlaAsnAlaLeuAlaIleGlnIlyLeuAla 177
DB 2535 GATATGAGTGCCTCAAGCTTTAGAGTGGCTAAAAATCTGCTGGAACGCGAGGTGGTGT 2594
QY 178 AspLysValSerPheGlnValAlaAspAlaLeuGlnIleProPheSerAspGlyGlnPhe 197
DB 2595 CATTTAATCAGTTTGAAGAGCAATGCAATGCAAACTTCCTATCAAGTGTAGTTT 2654
QY 198 AspLeuVal-----TPrSerMetClnSerGlyLuhMetProAspLys 212
DB 2655 GATATGTTTAATATGAGCATATGCTGACTATGCAAGCCGATCA-----GCTAAG 2705
QY 213 AlaLysPheValGlyLuhLeuAlaIleArgValAlaAlaProGlyAlaIleIleIle----- 230
DB 2706 AAAAAATGTATATGGAATATCTAAGGGTATTAACCTGAGAGTCTTCTCTTGACACAT 2765
```

```

QY 231 -----IleValThrTPrCysHis 236
DB 2766 GATGCTCTTAAAGAGCACTAAAGAGCTATTCAGACAGGAATATACAAAGCAATTCAT 2825
QY 237 ArgAspLeuGlyProAspGluGlnSerLeuHisProTPrGluGlnIleAspLeuLysLys 256
DB 2826 GTAATGTAGTCTTAACTCAAGAT-----GTTGGAGACAGGTGATGATGATCA 2879
QY 257 ---IleCysAsp 259
DB 2880 GGTATTGTGAT 2891

RESULT 13
US-08-576-626A-2
: Sequence 2, Application US/08576626A
: Patent No. 5998194
: GENERAL INFORMATION:
: APPLICANT: Summers, R.G.
: APPLICANT: Katz, L.
: APPLICANT: Donadio, S.
: APPLICANT: Slaver, M.J.
: TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR
: NUMBER OF SEQUENCES: 60
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Abbott Laboratories
: STREET: 100 Abbott Park Road
: CITY: Abbott Park
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60064-3500
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/576,626A
: FILING DATE: 21-DEC-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Dianne Casuto
: REGISTRATION NUMBER: P-40,943
: REFERENCE/DOCKET NUMBER: 5857.US.01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (847) 938-3137
: TELEFAX: (847) 938-2623
: TELEX:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 8051 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-576-626A-2

Alignment Scores:
Pred. No.: 0.000175 Length: 8051
Score: 131.00 Matches: 80
Percent Similarity: 33.23% Conserved: 32
Best Local Similarity: 23.74% Mismatches: 128
Query Match: 7.16% Indels: 97
DB: 2 Gaps: 19

US-09-857-613a-28 (1-350) x US-08-576-626A-2 (1-8051)
QY 18 PheArgSerClnSerProArg-----Thr 25
DB 2153 TGGGAGCGCCCTCACCAAGGTGTCGAGACACCCACCTACACCGAAGCGGTGAAGC 2212
```


GenCore version 5.1.4-P5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 19, 2003, 08:06:01 : Search time 88 Seconds
(without alignments) updates/sec
3084.279 Million cell

Title: US-09-857-613a-28
Perfect score: 1830
Sequence: 1 MATVRIPTISCHIHIFRS.....IEGYKDLIKFAIITCRKE 350

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 538826 seqs, 387737923 residues
Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09857613/rnunat.12032003.100038.6130/app-query.fasta.1.519
-DB=PublishedApplications_NA -QFWT=fastap -SUPTX=ripb -MINMATCH=0.1
-LOOPT=0 -LOOPT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09857613 -RCGN_1.1.53 -rnuat.12032003.100038.6130
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_MMAP -LARGOQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database: Published Applications_NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCIT_NEM_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEM_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEM_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCIT_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEM_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEM_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEM_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEM_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	386.5	21.1	252	US-09-878-574-695	Sequence 695, App
2	316	17.3	311	US-09-294-093B-3472	Sequence 3472, Ap
3	226	12.3	1080	US-09-938-842A-1815	Sequence 1815, Ap
4	226	12.3	1497	US-09-779-144A-6	Sequence 6, Appli

5	212.5	11.6	1086	9	US-09-938-842A-1856	Sequence 1856, Ap
6	203.5	11.1	1152	10	US-09-801-368-95	Sequence 95, Appl
7	203.5	11.1	1320	10	US-09-779-144A-1	Sequence 1, Appl1
8	203.5	11.1	1320	10	US-09-779-144A-5	Sequence 5, Appl1
9	196.5	10.7	1420	10	US-09-779-144A-3	Sequence 3, Appl1
10	156.5	8.6	3900	10	US-09-805-681-1	Sequence 1, Appl1
11	148.5	8.1	411	10	US-09-878-574-3629	Sequence 3629, Ap
12	141	7.7	753	9	US-10-260-877-53	Sequence 53, Appl1
13	140	7.7	1801	10	US-09-845-248-1	Sequence 1, Appl1
14	139	7.6	1353	10	US-09-738-626-636	Sequence 636, Ap
15	128	7.0	290	10	US-09-294-093B-5022	Sequence 5022, Ap
16	123	6.7	891	9	US-09-938-842A-1172	Sequence 1172, Ap
17	117.5	6.4	618	9	US-09-738-626-1423	Sequence 1423, Ap
18	114	6.2	261	10	US-09-938-626-984	Sequence 984, Ap
19	113	6.2	771	10	US-09-815-242-7986	Sequence 7986, Ap
20	113	6.1	4315	9	US-09-070-927A-1	Sequence 1, Appl1
21	112	6.1	753	10	US-09-738-626-2865	Sequence 2865, Ap
22	112	6.1	3309400	9	US-09-738-626-1499	Sequence 1, Appl1
23	110	6.0	654	9	US-09-938-842A-329	Sequence 1499, Ap
24	110	6.0	969	9	US-09-938-842A-1580	Sequence 329, Ap
25	104	5.7	5793	10	US-09-880-107-2109	Sequence 2109, Ap
26	101	5.5	699	10	US-09-974-300-1012	Sequence 1012, Ap
27	101	5.5	1406	10	US-09-939-980-1119	Sequence 119, Ap
28	99	5.4	726	10	US-09-815-242-8504	Sequence 8504, Ap
29	99	5.4	30365	10	US-09-825-414-1	Sequence 1, Appl1
30	98.5	5.4	702	10	US-09-815-242-4259	Sequence 4259, Ap
31	97.5	5.3	2313	9	US-09-938-842A-1580	Sequence 1580, Ap
32	97	5.3	801	12	US-10-007-693-50	Sequence 50, Appl
33	97	5.3	3309400	9	US-09-738-626-1	Sequence 1, Appl1
34	95.5	5.2	1836	9	US-09-938-842A-651	Sequence 651, App
35	95.5	5.2	3198	9	US-10-152-661-601	Sequence 601, App
36	95.5	5.2	3198	9	US-09-866-050A-601	Sequence 601, App
37	94.5	5.2	2040	9	US-09-866-050A-601	Sequence 1377, Ap
38	94.5	5.2	2040	9	US-09-738-626-1377	Sequence 7995, Ap
39	94	5.1	2340	10	US-09-815-242-7995	Sequence 1, Appl1
40	94	5.1	7596	10	US-09-728-952-1	Sequence 1281, Ap
41	93	5.1	1500	9	US-09-938-842A-1281	Sequence 2538, Ap
42	92.5	5.1	2811	9	US-09-938-842A-2538	Sequence 8247, Ap
43	92	5.0	2670	9	US-09-878-574-8247	Sequence 27, Appl
44	92	5.0	23907	9	US-09-927-827-27	Sequence 6, Appl1
45	92	5.0	24120	9	US-10-077-130-6	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-09-878-574-695
Sequence 695, Application US/09878574
Patent No. US20020110548A1
GENERAL INFORMATION:
APPLICANT: BYTUM, Joseph R.
APPLICANT: La Rosa, Thomas J.
APPLICANT: Thompson, Michael D.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(15401)B
CURRENT APPLICATION NUMBER: US/09/878,574
PRIOR FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/333,535
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO 695
LENGTH: 252
TYPE: DNA
ORGANISM: Glycine max
OTHER INFORMATION: Clone ID: LIB3028-047-Q1-B1-D12
US-09-878-574-695

Alignment Scores:
Pred. No.: 1.93e-36
Score: 386.50
Percent Similarity: 93.83%
Best Local Similarity: 90.12%
Length: 252
Matches: 73
Conservative: 3
Mismatch: 4

Query Match: 21.12% Indels: 1
DB: 10 Gaps: 1

US-09-857-613A-28 (1-350) x US-09-878-574-695 (1-252)

QY 66 GlnlysglyllealaglylphetrarpsgluserSerlyleuprglunslterpql 85
DB 9 CAGAGGGAATTCAGAGTTCACACAGATGCTGGCATTTGGAGACATTTGGGGC 68
QY 86 AsphismethishisgllyPheTytraspseraspserThrValSerleuSeraspHisarg 105
DB 69 GATCAGCAGCAGCAGCGCTTTATATGACCCGGATTCACCGCTTCTGTTGTCATCATGCG 128
QY 106 AlalaaglinleargmetileglnluserleuargpheaIaser---ValSerleu 124
DB 129 GCTGCTCAGATCCGATGATCCAGATCTTCTGTTGCTTCTTCTTCTGAGAAC 188
QY 125 ArgSerlystrprrpolsSerlyleValaspvalglycylglylSerSerarg 144
DB 189 CTTTCTAATATGCGCCAGAGATATAGTTGATGTTGGTGTGTCATTAAGGCGCAGCTCCAGA 248
QY 145 Tyr 145
DB 249 TAC 251

RESULT 2

US-09-294-093B-3472
Sequence 3472, Application US/09294093B
Patent No. US20010051335A1
GENERAL INFORMATION:
APPLICANT: Laligudi, Raghunath, V.
APPLICANT: Ito, Laura, Y.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
FILE REFERENCE: PL-0009 US
CURRENT FILING DATE: 1999-04-16
PRIOR FILING DATE: 60/082,567
NUMBER OF SEQ ID NOS: 6207
SOFTWARE: PERL Program
SEQ ID NO 3472
LENGTH: 311
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. US20010051335A1 700380165H1
NAME/KEY: unsure
LOCATION: 4, 43, 55, 96, 99, 127, 156
OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-3472

Alignment Scores:

Pred. No.: 4.31e-28 Length: 311
Score: 316.00 Matches: 66
Percent Similarity: 76.47% Conservative: 12
Best Local Similarity: 64.71% Mismatches: 23
Query Match: 17.27% Indels: 2
DB: 10 Gaps: 0

US-09-857-613A-28 (1-350) x US-09-294-093B-3472 (1-311)

QY 245 SerleuNisrptotrglinsrleuLyslylLecys-AspAlaTytrLeupr 264
DB 9 TCGSTAAACCCGATACCTGACCCCTCGAAGANATATGCGGANGCTACTACTCC 68
QY 264 AlaIatrcysSerThrSerAspTyValylsleuLeuInsrlSerleuInsprl 284
DB 69 GCACGTGCTCAGCTTCTGACTATGTNANNTGCCAAGTCACTCTCTCGAGATNT 128
QY 284 elysSerGluasprtrpserargpheaValaIarProhetrrpAlaValaIarSer 304
DB 304 AATTCGACAAAGACCCACA

DB 129 CACGACAGTCACTGCTCGAGACACTMGCCCGCTTTGGCCCGCCGATATAAATCAGC 188

QY 304 apheTrttrpLySgLyLeuSerleuSerleuSerGlylGlnlyThllySgLyAl 324

DB 189 GCTAATATGGAAGGCGCTTCACTCTGCTGACGACCGGATGGAACGATGAGAGCCG 248

QY 324 AleuAlaMetProLeuMeIleGlnlyTytrLysLysAspLeuIlelyPheAlaIle 344

DB 249 GATGCTGATGCGCTAATATGATCAGCGCTAACAAGAGGCGCTCATCAAT-TCACCATCAT 307

QY 344 eThr 345
DB 308 CACG 311

RESULT 3

US-09-938-842A-1815
Sequence 1815, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kieps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAIN
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
PRIOR FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 1815
LENGTH: 1080
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-1815

Alignment Scores:

Pred. No.: 8.89e-17 Length: 1080
Score: 226.00 Matches: 88
Percent Similarity: 42.21% Conservative: 61
Best Local Similarity: 24.93% Mismatches: 150
Query Match: 12.35% Indels: 54
DB: 9 Gaps: 13

US-09-857-613A-28 (1-350) x US-09-938-842A-1815 (1-1080)

QY 31 ValGlyProArgSerTrpAlaProIleArgAlaSerAlaIaSerSerGluArgGly 50
DB 70 CTAGTCCAGCAGAGAAAGGCAAGCAAGCAGCTGTGATCTCTCCGCG---GGCTCA 123
QY 51 lleValleuGlu-----GlnlyPro 57
DB 124 ATCTCCGCGAAGAAAGTCAAGACACTATACCAATACCTGTTCTTCTCCGAAACCA 183
QY 58 LysLys---AspAspLysLysLysLeuGlnLysGlylleAlaGluPheTytrAspGluSer 76
DB 184 AAGAGATTCGATACGCGCGAAGAAAGTCACTGCTGCTGACGCTTCTGCACTCTGCAATCTTCTGTC 243
QY 77 SerGlyLeuTrpGlnslleTrpGlyAspHisMetHis-----Gly 91
DB 244 ACTGATATCTACGAGTGGGATGGGACAACTTCTTCATTTCTCTCTCATGCTCCCTCGA 303
QY 92 PheTytrAspseraspserThrValSerleuSeraspHisargAlaIaGlnIleargMet 111
DB 304 AATTCGACAAAGACCCACA
QY 112 lleGlnIleSerleuargpheaIaSerValSerGlnGluArgSerlystrprrpSer 131
DB 131

```

Db 328 ATCCAGCAAGAAATG-----GCCGTCGATCTCATCAAGTGAACCGGGCAAAAG 378
Qy 132 ILeValaSpValGlyCysGlyIleGlySerSerArgTyrLeuAlaLysLysPheGly 151
    |||:||||| |||:||||| |||:|||||
Db 379 ATTCTGACCGCTGCTGGCGGTGGGCGCATGAGAGCCATGAGCCGCGCATTCAG 438
Qy 152 AlathSerValGlyIleThrLeuSerProValGlnAlaGlnAlaAsnAlaLeuAla 171
    |||:||||| |||:||||| |||:|||||
Db 439 GCCCAAGTCACTGCAATCATCATCAACAGTACCAAGTCAACAGCCAGCTTCACAAAC 498
Qy 172 AlalaGlnLeuAlaAlaApLysValSerPheGlnAlaAlaAspAlaGlnGlnPro 191
    |||:||||| |||:||||| |||:|||||
Db 499 AAGAAAGCTGAGTGTCTCTCTGCAACGCTGTTGCTTAACTTTTAAAGATGCCG 558
Qy 192 PheSerAspGlyGlnPheAspLeuValTyrSerMetGluSerGlyLysIleMetProAsp 211
    |||:||||| |||:||||| |||:|||||
Db 559 TTGATGATAAAGACGTTTGAAGGAGCTTACTCGATGAGAGCTTACGCTCCCTAAG 618
Qy 212 LysAlaLysPheValGlyGlnLeuAlaArgValAlaAlaProGlyAlaIleIleIle 231
    |||:||||| |||:||||| |||:|||||
Db 619 CTCGAAGAACTATATCTCGAGATCTTCAGATGATGAACACGAGATCTTGTCTGCC 678
Qy 232 ValThrTrpCysHisArgAspLeuGlyProAspGlnGlnSerLeuHisProTrpGlnGln 251
    |||:||||| |||:||||| |||:|||||
Db 679 TACCAATGGGTCACTGAAATATACAGAGCATGATCAAGAACAC-----AAG 729
Qy 252 AspLeuLeuLysLysIleCysAspAlaTyrTyrLeuProAlaTrpCysSerThrSerAsp 271
    |||:||||| |||:||||| |||:|||||
Db 730 GAGGTGATTCAGAGGATCGAGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 789
Qy 272 TyrValLysLeuLeuGlnSerLeuSerLeuGlnAspIleLysSerGluAspTrpSerArg 291
    |||:||||| |||:||||| |||:|||||
Db 790 ATGAGCCCTGACGCGGAGAAAGTGGGCTTGAAGTGAAGGAGGAGGAGGAGGAGGAG 849
Qy 292 PheValAla---ProPheTrpProAlaValIle-----ArgSerAlaPheThrTrpLys 308
    |||:||||| |||:||||| |||:|||||
Db 850 CCACCGCTAAACCGTGTGAACCGGTTAAAGATGGAAGATGCTTAT--TGAGGA 906
Qy 309 GLy-----LeuSerSerLeuLeuSerSerGlyGlnLysThrIleLysGlyAlaLeuAla 326
    |||:||||| |||:||||| |||:|||||
Db 907 AACCATTTGCTGTTGATTTCTTCTGCTATTTGGGTGCTGCTTAAAGGAGCTTGAT 966
Qy 327 MetProLeuMetIle-----GlnGlyTyrLysLysAspLeu 338
    |||:||||| |||:||||| |||:|||||
Db 967 GTTCATAGATGTTGTTAAGACTGCTGATTTATTTGACCAAGAGCTGGAGACTGGAATC 1026
Qy 339 IleLys---PheAlaIleIleThrCysArgLysProGlu 350
    |||:||||| |||:||||| |||:|||||
Db 1027 TTCTCTCCGATGATATGATTTCTCTGTAGAAACCAAGAG 1065

```

```

RESULT 4
US-09-779-144A-6
; Sequence 6, Application US/09779144A
; Patent No. US20020148006A1
; GENERAL INFORMATION:
; APPLICANT: Nes, David W.
; TITLE OF INVENTION: Transgenic Plants with Modified Sterol Compositions
; FILE REFERENCE: 11899.0198, DIVUS01, MOBF.198
; CURRENT APPLICATION NUMBER: US/09/779,144A
; PRIORITY FILING DATE: 2001-02-08
; PRIOR FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 1497
; TYPE: DNA
; ORGANISM: Zea mays
US-09-779-144A-6

```

```

Alignment Scores: 1.47e-16 Length: 1497
Pred. No.: 226.00 Matches: 81
Score:

```

```

Percent Similarity: 38.77% Conservative: 45
Best Local Similarity: 24.92% Mismatches: 121
Query Match: 12.35% Indels: 78
DB: 10 Gaps: 10

US-09-857-613a-28 (1-350) x US-09-779-144A-6 (1-1497)
Qy 69 ILeAlaGlnPheThrAspGluSerSerGlyLeuTrpGlnAlaIleTrpGlyAspHisMet 88
    |||:||||| |||:||||| |||:|||||
Db 216 GTTATATAATATCATATGATCTTCCACATGCTTCTATAGATATGTTGGTGGATGATCTTC 275
Qy 89 HisIleGlyPheThrAspSerAspSerThrValSerLeuSerAspHisArgAlaAlaGln 108
    |||:||||| |||:||||| |||:|||||
Db 276 CACTTTCCTCAAGATGGAATGAGAAATCCCTTACGTGAACATCAACGACATGACAT 335
Qy 109 IleArgMetIleGlnGlnSerLeuArgPheAlaSerValSerGlnGlnArgSerLysTrp 128
    |||:||||| |||:||||| |||:|||||
Db 336 TTCTTGGCCCTGCAACTGTTGTTG-----AAACCA 365
Qy 129 ProLysSerIleValaSpValGlyCysGlyIleGlyLysSerArgTyrLeuAlaLys 148
    |||:||||| |||:||||| |||:|||||
Db 366 GCAATGAAAGCTTTTATGATGTGGCTGTGATGAGTGGACCATGAGAAATGCA--- 422
Qy 149 LysPheGlyAlaThrSerVal---GlyIleThrLeuSerProValGlnAlaGlnAla 167
    |||:||||| |||:||||| |||:|||||
Db 423 AGATTAGCTCAACTTCACTTACCGGATTTGAATCAACGATACCAAGATACAGGGGA 482
Qy 168 AsnAlaAlaAlaAlaGlnGlyLeuAlaAspLysValSerPheGlnValAlaAlaAla 187
    |||:||||| |||:||||| |||:|||||
Db 483 AAGAGCTCAACCGTTTACGAGAAATTAAGTGAACATGTGATTTGTCAAGCGGACTTC 542
Qy 188 LeuGlnGlnProPheSerAspGlyGlnPheAspLeuValTyrSerMetGluSerGlyLys 207
    |||:||||| |||:||||| |||:|||||
Db 543 ATGAAGATGCCCTTGCATGACACACTTTGATGCTTTTACCGCATTTAGAGGACCATGCT 602
Qy 208 HisMetProAspLysAlaLysPheValGlyGlnLeuAlaArgValAlaAlaProGlyAla 227
    |||:||||| |||:||||| |||:|||||
Db 603 CATGACCTGATCCAGTGTGTTGCTACAGAGATATATCGTGTGAAGCTGGCCAG 662
Qy 228 IleIleIleIleValaThrTrpCys-----His 236
    |||:||||| |||:||||| |||:|||||
Db 663 TGCCTTGGCGGTGACAGAGTGGTGGCTTACGATCACTATGATCTCAACATGCAACCCAC 722
Qy 237 Arg-----AspLeuGly-----ProAspGlnGlnSerLeuHis 247
    |||:||||| |||:||||| |||:|||||
Db 723 AAAAGATCAAGATGAAATGAGCTTGGCAATGSCCTGCAGATATCAAGAGCCTGG 782
Qy 248 Pro-----TrpGlnGlnAspLeu 254
    |||:||||| |||:||||| |||:|||||
Db 783 CAATGCTCCGCGAGTAAAGACCGCGGTTTGAGGTTGTTGGATGAAGATCTTCT 842
Qy 255 LysLysIleCysAspAlaTyrTyrLeuProAlaTrpCysSer----- 268
    |||:||||| |||:||||| |||:|||||
Db 843 GAAATTTCTCCCTTGGCTTGGTACTTGGCTTGATTCACACCGCATTCCTCCGTGATGTC 902
Qy 269 -----ThrSerAspTyrValLysLeuLeuGlnSer 278
    |||:||||| |||:||||| |||:|||||
Db 903 TTCCGTTTGACCTCTGTGGAGCATGATTTACCCGACAAATGCTAAGAGCCCTGAGATAC 962
Qy 279 LeuSerLeuGlnAspIleLysSerGlnAspTrpSerArgPheValAlaProPheThrPro 298
    |||:||||| |||:||||| |||:|||||
Db 963 GTTGCTCTGCTCCGACAGGAGTGAAGAGGCTCTACTTTCCTGAG----- 1010
Qy 299 AlaValIleArgSerAlaPheThrTrpLysGlyLeuSerSerLeuLeuSerSerGlyLys 318
    |||:||||| |||:||||| |||:|||||
Db 1011 -----AAGCTGCAAGAGGCTGTGAGAGGCGGAAAG 1043
Qy 319 LysThrIleLysGlyAlaLeuAlaMetProLeuMetIleGlnGlyTyrLysLysAspLeu 338
    |||:||||| |||:||||| |||:|||||
Db 1044 AAGGAGATC-----TTCAAGCCATGATCTTCTTTTGTTCGGAAGCCTCTT 1091
Qy 339 IleLysPheAlaIle 343
    |||:||||| |||:||||| |||:|||||

```

Db 1092 CTGGAATGAGCTCTT 1106

RESULT 5
US-09-938-842A-1856
Sequence 1856, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 1856
LENGTH: 1086
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-1856

Alignment Scores:
Pred. No.: 3,35e-15 Length: 1086
Score: 212.50 Matches: 80
Percent Similarity: 42.54% Conservatively: 54
Best Local Similarity: 25.40% Mismatches: 144
Query Match: 11.61% Indels: 37
DB: Gaps: 11

US-09-857-613A-28 (1-350) x US-09-938-842A-1856 (1-1086)

QY 55 GlnLysProLysLys---AspAspLysLysLysLeuGlnLysGlyLeuAlaGluPheTyr 73
Db 175 CGCGCTCCAAAGAAATGCAACCGCGCGGAAAGTTCACAGCTTCGCGACACATTTCAC 234
QY 74 AspGluSerSerGlyLeuPheGlnLysLysLysLysLysLysLysLysLysLysLys 93
Db 235 AATCTGCTCCACCGCATATATGAGTGGGATGGGACATCTCCAC-----TCTCA 288
QY 94 AspSerAspSerThrValSerLeuSerAsp-----HisArgAlaAlaGlnIle 109
Db 289 CCATCATTCCTCCGAAATATCTCACAAAGACGCGCTCCACGAAGAGATGCGGTA 348
QY 110 ArgMetIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 129
Db 349 GATCTGATCCAA-----GTCAACCTGCT 372
QY 130 LysSerIleValAspValGlyCysGlyIleGlyGlySerArgTyrLeuAlaLysLys 149
Db 373 CAAAGATCTTACGATGCGGATGCGGCTCCGATCCGACGATGCTTCAC 432
QY 150 PheGlyAlaThrSerValGlyIleThrLeuSerProValGlnAlaGlnAlaAsnAla 169
Db 433 TCGGAGCTAAGTACGATGCGGATTAACGATACGATGATGATGATGATGATGATGAT 492
QY 170 LeuAlaAlaAlaGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGln 189
Db 493 CACATTAAGAAAGCTGCTGACGCGCTTCGAGGCTGCTGTTACTCTCCAG 552
QY 190 GlnProPheSerAspGlyGlnPheAspLeuValTyrSerMetGluSerGlyGluHisMet 209
Db 553 ATGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 612
QY 210 ProAspLysAlaLysPheValGlyLeuAlaArgValAlaAlaProGlyAlaIleIle 229

Db 613 CCGAAGCTGAGAAGATGTACGACAGATCTACAGGGTGTGAACCGGATCTATGTAT 672
QY 230 IleIleValIleThrPheCysHisArgAspLeuLysProAspGluGlnSerLeuHisProThr 249
Db 673 GTGTCTGACGATGGGTTCAGACGAGAAATTTAAGCGGAGATGACGAACAC----- 726
QY 250 GluGlnAspLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 269
Db 727 ---GTGAGGATATCCAGGATTCAGAGCGGATTCAGAGCGGATTCAGAGCGGATTCAG 783
QY 270 SerAspTyrValLysLeuGlnSerLeuSerLeuGlnAspIleLysSerGlnAspThr 289
Db 784 GTGATATATAGCTGACAGCGCTTAAAGGTTGGTTGATGATGATGATGATGATGATGAT 843
QY 290 SerArgPheValAla---ProPheThrProAlaValIle-----ArgSerAlaPheThr 306
Db 844 GCGAGTCCACCGCTGAGCGCGGTGAGCTAGCGCTTAAGATGCGTGAAGCGTCTAT--- 900
QY 307 TrpLysGly-----LeuSerLeuLeuSerSerGlyGlnLysThrIleLysGlyAla 324
Db 901 TGGAGGATACATTTGTTGTTGATTTTGTACGCGGTTGAGGTTGCTCTTAAGCAACT 960
QY 325 LeuAlaMetProLeuMetIle-----GluGlyTyrLysLys 336
Db 961 GTTGATGTTCAATGATGATGTTGTTAAGCTGCTGATTTATTTGACAGAGAGTGAAC 1020
QY 337 AspLeuIleLys---PheAlaIleIleThrCysArgLysProGlu 350
Db 1021 GGAATATCTCTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1065

RESULT 6
US-09-801-368-95
Sequence 95, Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Calli, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: No. US20020128250A1man, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801,368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 95
LENGTH: 1152
TYPE: DNA
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-95

Alignment Scores:
Pred. No.: 4.09e-14 Length: 1152
Score: 203.50 Matches: 78
Percent Similarity: 37.99% Conservatively: 58
Best Local Similarity: 21.79% Mismatches: 149
Query Match: 11.12% Indels: 73
DB: Gaps: 12

US-09-857-613A-28 (1-350) x US-09-801-368-95 (1-1152)

```

Qy 42 SerAlaIaSerSerGluArgGlyGluIleValLeuGluGlnLysProLysLysAsp 61
Db 106 AACTCTGCCCCAAAGAAAGCCGCTCAGAACTTGGCAATTTGGATGATGAGAACGAT 165
Qy 62 LysLysLysLeuGlnLysGlyIleAlaGluPheTyrAspGluSerSerGlyLeuTyrP 81
Db 166 AAGATGCGCAAGAAAGCGTCTT---GAGATATATATGAAGCCACATCTCTACTAT 222
Qy 82 AsnIle-----TrrGlyAspHisMetHisGly---Phe 92
Db 223 AACCTCGTACAGATTCTATGAATATGTTGGGCTTCTTCCATTTCCAGACAGATT 282
Qy 93 TyrAspSerAspSer-----ThyAlaSerLeuSerAspHisArgAlaGlnIle 109
Db 283 TATTAAGGTGAGAGATTGCTGCTGCTCCGCTACAGACATGACATTTATTTAGCTTACAG 342
Qy 110 ArgMetIleGlnGluSerLeuArgPheAlaSerValSerGluGluArgSerLysTyrPro 129
Db 343 GCTGATTTCAAGAGCC----- 360
Qy 130 LysSerIleValAspValGlyCysGlyIleGlyLysSerArgTyrLeuAlaLysLys 149
Db 361 GATTAGTTCGACGCTGTTGCTGCTGCTGCGGCCAGCAGAGAGATTCAGAGATT 420
Qy 150 PheGlyAlaThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaAsnAla 169
Db 421 ACCGCTGTACAGTCATCGTCTGCTTAACAATACATTCCTCAATTTCCAAAGCCAAATAT 480
Qy 170 LeuAlaAlaIaGlnGlyLeuAlaAspLysValSerPheGlnValAlaAlaLeuGln 189
Db 481 TACGTTAAATACAAATTTGATGATGACCAATGACATTTGTAAGGATTTATTCATGAAA 540
Qy 190 GlnProPheSerAspGlyGlnPheAspLeuValTyrPserMetGluSerGlyLysHisMet 209
Db 541 ATGATTTTCGAGAAACACTTTCGACAAAGTTTATGCAATTTGACAGCCATGTCACCT 600
Qy 210 ProAspLysAlaLysPheValGlyGluLeuAlaArgValAlaAlaProGlyAlaIleIle 229
Db 601 CCAAAATTAAGAGGTGTATACAGCAATCTACAGATTTTGAACCGGTGTCACCT 660
Qy 230 IleIleValIleThrPrcYshisArgAspLeuGlyProAspGluGlnSerLeuHisProTyr 249
Db 661 GCTGTTTACGAATGGGTATGACTGATTAATATGACGAAACATCTGACAT----- 714
Qy 250 GluGlnAspLeuLeuLysIleCysAspAlaTyrTyr-----Leu 263
Db 715 -----AGAAAGATC-----GCTATGAATTTGAACCTAGTGTATGCTATC 753
Qy 264 ProAlaIleTyrCysSerThrSerAspTyrValLysLeuLeuGlnSerLeuSerLeuGlnAsp 283
Db 754 CCAAAAGATTTCCATGCTGACGCTGCTAGAGAAAGCATTTGAACATCTGGTTTCCAGATC 813
Qy 284 IleLysSerGluAsp-----TyrPserArgPheValAla 294
Db 814 CTCGTTAGCAAGAGCTGCGGCGACATGATGTAATCCCTTGGATTAACCATTA 873
Qy 295 ProPheTrrProAlaVal-----IleArgSerAlaPhe 305
Db 874 GGTGAGTGAAGTACGTTCAAAACTTACGTAATTTGCCACATTTTTCAGAACTTCTTAC 933
Qy 306 ThrTrrPlyLysLeuSerSer-----LeuLeuSer 315
Db 934 TTGGGTAGCAATTTACTACAGCAATGTTACTGTAATGAAATTTAGTCTTACGCCCA 993
Qy 316 SerGlyGlnLysThrIleLysGlyAlaLeu-----AlaMetProLeuMetIleGln 332
Db 994 GAAGGTTCACAGAGATTAATCTGCTGCTAGAAAATGCGCGGTTGGTTACTTGGCGGT 1053
Qy 333 GlyTrrLysLysAspLeuIleLysPheAlaIleIleThrCysArgLysProGlu 350
Db 1054 GGTAAAGTCCAAAGTATTAATCACTCCAAATGATGCTTTTCGTCGTAGAGAGCCAGAA 1107

```

```

RESULT 7
US-09-779-144A-1
; Sequence 1, Application US/0979144A
; Patent No. US20020148006A1
; GENERAL INFORMATION:
; APPLICANT: Nes, David W.
; TITLE OF INVENTION: Transgenic Plants with Modified Sterol Compositions
; FILE REFERENCE: 11899,0198, DVUS01 MOBT:198
; CURRENT APPLICATION NUMBER: US/09/779,144A
; PRIOR FILING DATE: 2001-02-08
; PRIOR APPLICATION NUMBER: US 09/106,926
; PRIOR FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1320
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-779-144A-1

```

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
5.05e-14	203.50	37.99%	21.79%	11.12%	1320	78	58	149	73	12

US-09-857-613a-28 (1-350) x US-09-779-144A-1 (1-1320)

```

Qy 42 SerAlaIaSerSerGluArgGlyGluIleValLeuGluGlnLysProLysLysAsp 61
Db 177 AACTCTGCCCCAAAGAAAGCCGCTCAGAACTTGGCAATTTGGATGATGAGAACGAT 236
Qy 62 LysLysLysLeuGlnLysGlyIleAlaGluPheTyrAspGluSerSerGlyLeuTyrP 81
Db 237 AAGATGCGCAAGAAAGCGTCTT---GAGATATATATGAAGCCACATCTCTACTAT 293
Qy 82 AsnIle-----TrrGlyAspHisMetHisGly---Phe 92
Db 294 AACGTCGTACAGATTCTTATGAATATGTTGGGCTTCTTCCATTTTCCAGACATTT 353
Qy 93 TyrAspSerAspSer-----ThyAlaSerLeuSerAspHisArgAlaGlnIle 109
Db 354 TATAAAGGTGAGAGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 413
Qy 110 ArgMetIleGlnGluSerLeuArgPheAlaSerValSerGluGluArgSerLysTyrPro 129
Db 414 GCTGATTTCAAGAGCC----- 431
Qy 130 LysSerIleValAspValGlyCysGlyIleGlyLysSerArgTyrLeuAlaLysLys 149
Db 432 GATTAGTTCGACGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 491
Qy 150 PheGlyAlaThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaAsnAla 169
Db 492 ACCGCTTGAACGTCATCGCTTAACAATACATTAACCAATTTGCAAGCCAAATAT 551
Qy 170 LeuAlaAlaIaGlnGlyLeuAlaAspLysValSerPheGlnValAlaAlaLeuGln 189
Db 552 TACGTTAAATAATACAAATTTGATGATGACCAAAATGCTTGTAAAGGCTTATTCATGAAA 611
Qy 190 GlnProPheSerAspGlyGlnPheAspLeuValTyrPserMetGluSerGlyLysHisMet 209
Db 612 ATGATTTTCGAGAAACACTTTCGACAAAGTTTATGCAATTTAGAGGCCATGTCACGCT 671
Qy 210 ProAspLysAlaLysPheValGlyGluLeuAlaArgValAlaAlaProGlyAlaIleIle 229
Db 672 CCAAAATTAAGAGGTGTATACAGCGCAATCTACAGATTTTGAACCGGTGTCACCTT 731
Qy 230 IleIleValIleThrPrcYshisArgAspLeuGlyProAspGluGlnSerLeuHisProTyr 249
Db 732 GCTGTTTACGAATGGGTATGACTGATTAATATGACGAAACATCTGACAT----- 785

```

```

OY 250 GIUGLINSPLLEULYSILIECYSPALATYTYR-----Leu 263
Db 786 -----AGAAAGATC-----GCTATGAATTCAGTACGTGATG 824
OY 264 PROALATPCYSSETHSERASPTYRVALLYSLEULEGINSERLEUSERLEUGLINSASP 283
Db 825 CCAAAGATTCCTCAGTGCAGCTGCTAGCAAGCATGTGAAGAACTGCTGTTCGAAGTC 884
OY 284 ILELYSSEGLUASP-----TRPSERARGPHEVALALA 294
Db 885 CTCGTTAGCGAAGACCTGCGGAGCAATGATGATAATCCCTTGATTAACCATTA 944
OY 295 PROPHETIPROALAVALE-----LEARGSERALAPHE 305
Db 945 GGTGAGTGGAGTACGCTTCAAAACTAGCTAATTTGCCACATTTTTCAGAACTTCTAC 1004
OY 306 THRTPYSGLYLEUSER-----LEULEUSER 315
Db 1005 TTGGGTAGACAAATTACTACAGCAATGTTACTGTATGAGAAAATTAGCTTACGCCCA 1064
OY 316 SERGLYGLINLTHRLIELYSGLYALALEU-----ALAMETPROLEUMETILEGLU 332
Db 1065 GAAGGTTCCAGAGAACTTCTGCTGTAGAAAATCTCGCTGGTTAGTTCGCCGT 1124
OY 333 GLTYRILYSLSASPLEULIELYSPHEALALELLELTHRCYSARGLYSPROGLU 350
Db 1125 GGTAACTCCAAAGTATTCACTCCATGATGCTTTTCTGCTAGAGAGCCAGAA 1178

RESULT 8
US-09-779-144A-5
: Sequence 5, Application US/09779144A
: Patent No. US20020148006A1
: GENERAL INFORMATION:
: APPLICANT: Nes, David W.
: TITLE OF INVENTION: Transgenic Plants with Modified Sterol Compositions
: FILE REFERENCE: 11899, 0198, DVS01, MOBT, 198
: CURRENT APPLICATION NUMBER: US/09/779,144A
: CURRENT FILING DATE: 2001-02-08
: PRIOR APPLICATION NUMBER: US 09/106,926
: PRIOR FILING DATE: 1998-06-29
: NUMBER OF SEQ ID NOS: 16
: SOFTWARE: Patentin version 3.0
: SEQ ID NO 5
: LENGTH: 1320
: TYPE: DNA
: ORGANISM: Saccharomyces cerevisiae
: US-09-779-144A-5

Alignment Scores:
Pred. No.: 5 05e-14 Length: 1320
Score: 203.50 Matches: 78
Percent Similarity: 37.99% Conservative: 58
Best Local Similarity: 21.79% Mismatches: 149
Query Match: 11.12% Indels: 73
DB: 10 Gaps: 12

US-09-857-613A-28 (1-350) x US-09-779-144A-5 (1-1320)
OY 42 SERALALASERSEGLUARGLYGLULELEUGLINSERPROLYSLSASPSASP 61
Db 177 AACTCTGCCCAAAAGACCGCTTCAGAACTTACGAAATTTGGATGATGATGATGAT 236
OY 62 LYSLSYLSLEUGLINSGLYILEALAGLUHERYASPSGLUSERSEGLYNEUTRPGLU 81
Db 237 AAAGATGCCAGAGAACGCTCTT---GAGATTAATAATGAAGCCACACATTCCTACAT 293
OY 82 ASUNLE-----TRPGYASPHISMETHISISGLY---PHE 92
Db 294 AACGTCGTTACAGATTTTATGAAATATGCTTGCGGCTCTCTTCATTTTCAGCAGATT 353
OY 93 TYRASPSEASPSE-----THRVASERLEUSERASPHISARGALALAGLNULE 109

```

```

Db 354 TATTAAGTGACAGTTTCCTGCTCGATGACAGACATGAACTATTATTACCTACAG 413
OY 110 ARGMETILEGLINSERLEUARGPHEALASERVALSERGLUARGSERLYSTRPPO 129
Db 414 GCTGCTATCCAAAGAGCC-----
OY 130 LYSSERIIEVALASPAVALCYSGLYILEGLYGLYSERSEARGTYRILEUALAYSLS 149
Db 432 GATTTAGTCTCGACCTGCTGTGCTGTGGGGCCGAGCAAGACAGATTCGAATTT 491
OY 150 PHEGLVALATHSERVALGLYILIEHRLHUSERPROVALIGLALAGLNRGALAAANALA 169
Db 492 ACCGTTGTAGCTCATGCTGTAAACAATAACGATTTACCAATTGCCACAGCAAAATAT 551
OY 170 LEUALALALAGLINSLEUALAASPLYSVALSERPHEGLINVALALASPALALEUGL 189
Db 552 TACGCTAAATAATACAAATTTAGCTACCAATAGCACTTTGTAAGGGTATTTACATGAA 611
OY 190 GLNPROPHESERASPOLYGLINPHEASPLEUVALTRPSETHECGLUSERGLYUNHISMET 209
Db 612 ATGCAATTTCCAGAAACAACCTTTCGACAAAGTTTATGCAATTTAGGCAATCTCACGCT 671
OY 210 PROASPLYSALALYSRPHLEVALIGLYLULEUALARGVALALALALAPROGLYALALE 229
Db 672 CCAAAATTAGACGTGATATACAGCAATCTACAGATTTTGAACCGGTGCTACCTTT 721
OY 230 ILELEVALATHRTPCYSHISARGASPLEUGLYPROASPLUGLINSERLEUHSPTOTIP 249
Db 732 GCTGTTACGAATGGTAATGATGATGAATAATATGACGAAACAATCCTGTAACAT 785
OY 250 GIUGLINSPLLEULYSILIECYSPALATYTYR-----Leu 263
Db 786 -----AGAAAGATC-----GCTATGAATTCAGTACGTGATGATG 824
OY 264 PROALATPCYSSETHSERASPTYRVALLYSLEULEGINSERLEUSERLEUGLINSASP 283
Db 825 CCAAAGATTCCTCAGTGCAGCTGCTAGCAAGCATGTGAAGAACTGCTGTTCGAAGTC 884
OY 284 ILELYSSEGLUASP-----TRPSERARGPHEVALALA 294
Db 885 CTCGTTAGCGAAGACCTGCGGAGCAATGATGATAATCCCTTGATTAACCATTA 944
OY 295 PROPHETIPROALAVALE-----LEARGSERALAPHE 305
Db 945 GGTGAGTGGAGAGTACTTCAAACTTAGCTAATTTGGCCACATTTTTCAGAACTTCTAC 1004
OY 306 THRTPYSGLYLEUSER-----LEULEUSER 315
Db 1005 TTGGGTAGACAAATTACTACAGCAATGTTACTGTATGAGAAAATTAGCTTACGCCCA 1064
OY 316 SERGLYGLINLTHRLIELYSGLYALALEU-----ALAMETPROLEUMETILEGLU 332
Db 1065 GAAGGTTCCAGAGAACTTACTGCTGTAGAAAATCTCGCTGGTTAGTTCGCCGT 1124
OY 333 GLTYRILYSLSASPLEULIELYSPHEALALELLELTHRCYSARGLYSPROGLU 350
Db 1125 GGTAACTCCAAAGTATTCACTCCATGATGCTTTTCTGCTAGAGAGCCAGAA 1178

RESULT 9
US-09-779-144A-3
: Sequence 3, Application US/09779144A
: Patent No. US20020148006A1
: GENERAL INFORMATION:
: APPLICANT: Nes, David W.
: TITLE OF INVENTION: Transgenic Plants with Modified Sterol Compositions
: FILE REFERENCE: 11899, 0198, DVS01, MOBT, 198
: CURRENT APPLICATION NUMBER: US/09/779,144A
: CURRENT FILING DATE: 2001-02-08
: PRIOR APPLICATION NUMBER: US 09/106,926
: PRIOR FILING DATE: 1998-06-29
: NUMBER OF SEQ ID NOS: 16
: SOFTWARE: Patentin version 3.0
: SEQ ID NO 3

```

LENGTH: 1420
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-779-144A-3

Alignment Scores:

Pred. No.:	3 7e-13	Length:	1420
Score:	196.50	Matches:	81
Percent Similarity:	42.86%	Conservative:	54
Best Local Similarity:	25.71%	Mismatches:	143
Query Match:	10.74%	Indels:	38
DB:	10	Gaps:	11

US-09-857-613a-28 (1-350) x US-09-779-144A-3 (1-1420)

QY 55 GlnIysProLysLys---AspAspLysLysLysLeuGlnIysGlyIleAlaGluPheTyr 73
DB 217 CGCGCTCCAAAGAAATCGAAACCGCGGAGCAAGTTCCAGACTTCGTCGACACATCTAC 276
QY 74 AspGluSerSerGlyLeuTrrpGluAsnIleTrrpGlyAspHisMetHisGlyPheTyr 93
DB 277 AATCTGCTCCACACATATCGATGGGATGGGACATCTTCAC-----TTCCTCA 330
QY 94 AspSerAspSerThrValSerLeuSerAsp-----HisArgAlaAlaGlnIle 109
DB 331 CCATCAATCCCCGAAATATTCACAAAGACCCACGCCCTCCACGAGACATGGCGGTA 390
QY 110 ArgMetIleGlnIysLeuArgPheAlaSerValSerGluArgSerLysTrrpPro 129
DB 391 GATCTGATCCA-----GTCAACCTGGT 414
QY 130 LysSerIleValAspValGlyCysGlyIleGlyLysSerArgTrrpLeuAlaLysLys 149
DB 415 CAAAGATCTTAACGCGGATGCGGTGCGGTCGATCGCAGCGCATTCCTAC 474
QY 150 PheGlyIleThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaAsnAla 169
DB 475 TCCGAGCAACGTA-GTCGGGATTACAATAACAGATATCGTGACAGACGCTGCTC 533
QY 170 LeuAlaAlaIleGlnIysLeuAlaAspLysValSerPheGlnValAlaAspAlaLeuGln 189
DB 534 CACAAATGAAGAGCTGCTCGACGCCCTTGCAGAGTCGTGTGGTAACCTCTCCAG 593
QY 190 GlnProPheSerAspGlyGlnPheAspLeuValTrrpSerMetGluSerGlyGlnHisMet 209
DB 594 ATGCCGTTTCATGACACAGTTCGACGAGCTTATTCATCGAAGCCAGCTGTCACCG 653
QY 210 ProAspLysAlaLysPheValGlyGlnLeuAlaArgValAlaIleProGlyAlaIleLeu 229
DB 654 CCGAAGCTGGAAGAGTGTACGACAGATCTACAGGGGTGTGAACCCGGAATCTATGTAT 713
QY 230 IleIleValThrTrrpCysHisArgAspLeuGlyProAspGluGlnSerLeuHisProTrrp 249
DB 714 GTGCGTACAGGTGTTACGACGAGAAATTTAAGCGGAGTACGACGACAC----- 767
QY 250 GlnGlnAspLeuLeuLysLysIleCysAspAlaTrrpTrrpLeuProAlaTrrpCysSerThr 269
DB 768 ---GTGAGGATATCCAAAGGATTGACAGAGCGATTCAGCGGCTTACGCGCTTAC 824
QY 270 SerAspLysValLysLeuGlnSerLeuSerLeuGlnAspIleLysSerGluAspTrrp 289
DB 825 GTGATATAGCTGAGACGCTTAAGAGGTTGGTTTGAATAGTACAGAGAGAGGATCTG 884
QY 290 SerArgPheValAla---ProPheTrrpProAlaValIle---ArgSerAlaPheThr 306
DB 885 CGGAGTCCACCGGCTGAGCGCTGCTAGCGTACGCTTAGATGGGTAGCGCTTCTAT--- 941
QY 307 TrrpLysGly-----LeuSerSerLeuLeuSerSerGlyGlnLysThrIleLysGlyAla 324
DB 942 TGGAGGATTCACATTTGCTGCTCAGATTTTGTACAGCGTTGAGTTGCTCTAAAGGACT 1001
QY 325 LeuAlaMetProLeuMetIle-----GluGlyTrrpLysLys 336

DB 1002 GTTCATGTTTCATGAGATCTTGTAAAGACTGATGTTTGAACGAGAGGTGAAC 1061
QY 337 AspLeuIleLys---PheAlaIleIleThrCysArgLysProGlu 350
DB 1062 GGAATATTTCTCTCCGATGATGATGATTTCTGCGAAGAACCGGAG 1106

RESULT 10

US-09-805-681-1

Sequence 1, Application US/09805681
Patent No. US20020143163A1
GENERAL INFORMATION:
APPLICANT: INOUE, MASAYORI
APPLICANT: PHADTARE, SANGITA
APPLICANT: YAMAKA, KUNITOSHI
APPLICANT: KATO, IKUNOSHIN
TITLE OF INVENTION: ANTIBACTERIAL ACTIVITY OF 4,5 DIHYDROXY-2-CYCLOPENTAN-1-ONE
TITLE OF INVENTION: (DHCP) AND CLONING A GENE CONFERRING DHCP RESISTANCE IN
FILE REFERENCE: 1137-P-00
CURRENT APPLICATION NUMBER: US/09/805,681
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/228,727
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 3900
TYPE: DNA
ORGANISM: Escherichia coli
US-09-805-681-1

Alignment Scores:

Pred. No.:	8e-08	Length:	3900
Score:	156.50	Matches:	69
Percent Similarity:	41.45%	Conservative:	45
Best Local Similarity:	25.09%	Mismatches:	105
Query Match:	8.55%	Indels:	56
DB:	10	Gaps:	14

US-09-857-613a-28 (1-350) x US-09-805-681-1 (1-3900)

QY 71 GluPheTrrpAspGluSerSerGlyLeuTrrpGluAsnIleTrrpGlyAspHisMetHis 90
DB 2100 GAGCATTTACGATTTGGGATATGCTTTCAGCCGATGATCCCTTCATGCAATAT 2159
QY 91 Gly-----PheTrrpAspSerAspSerThrValSerLeuSerAspHisArgAla 107
DB 2160 TCCCTCGCTTACTGGAAGATCCGAT-----AATCTGATCTGCCACGACGCG 2210
QY 108 GlnIleArgMetIleGlnIysLeuArgPheAlaSerValSerGluGlnArgSerLys 127
DB 2211 AAGCTCAAAATATGTAAGAAATTCGAGTTA-----AAA 2246
QY 128 TrrpProLysSerIleValAspValGlyCysGlyIleGlyLysSerArgTrrpLeuAla 147
DB 2247 CCAGGATCCGCTACTGATATTTGGCTCGGCTGGGGGAGCTGACACATGAGCA 2306
QY 148 LysLysPheGlyAlaThrSerValGlyIleThrLeuSerProValGlnAlaArgAla 167
DB 2307 TCTAATTATGAGTAAAGCTGTGGGCTCACCATTTCTGCGAAGCAAGAAATGGCT 2366
QY 168 AsnAlaLeuAlaAlaGlnIysLeuAlaAspLysValSerPheGlnValAlaAspAla 187
DB 2367 CAG-----GAAGCTGTGAAGCCCTGATGACCAT----- 2399
QY 188 LeuGlnGlnProPheSerAsp-----GlyGlnPheAspLeuValTrrpSerMetGluSer 205
DB 2400 TTGCTGCAAGATTTATCTGACCTGAACGACGACGTTGATGATGTTCTTGCGGATG 2459
QY 206 GlyGlnHisMetProAspLys-----AlaLysPheValGlyGlnLeuAlaArgValAla 223
DB 2460 TTGAGACAGCTGAGACCGAAATTAACATACCTATTTTGGCGTGGTGTGATGTAATTTG 2519

Qy	224	AlaproglylailellellellelleValITPTTCpysHisIargaspLeuGlyProaspIu	243
		:::::	::: :::
Db	2520	AAACCGAAGCATATTCCTCGCTCCACTACT	ATCGCTCGAAAAA 2564
Qy	244	GlnSerLeuHis-----ProTrpGluGlnaspLeuLeuLysLysIleOysaspAlaTyr	261
		:::::	::: :::
Db	2565	ACCGATCTGCATCTGCATCTCCCTCG-----ATTAAATAAATAT 2600	
Qy	262	TyrlLeuProAlaTrpCysSerThrSeraspTyrValLysLeuLeuGlnSerLeuSerLeu	281
			:::::
Db	2601	ATTTCCTCGAAGCGTTCCCTCGCCCTCT-----GTACGCAATGTGCATGCGCAGCGCA	2654
Qy	282	GlnaspIleLysSerGluaspTrpSerAlaGpHeValAlaProPheTrpProAlaValIle	301
			::: :::
Db	2655	CCCCACTTGTGATGAGCAAGCATGGCAATACCTTCGGTGCATGATTAC-----	2699
Qy	302	ArgSerAlaPheThr-TripLysLysLeuSerSerLeuLeuSerSerGlyGlnLys-----	319
		::::: :::::	
Db	2700	---GATACACTACGTTGAGCGGGGTGTGAACAGATTCTCTCGCCGCAATGGCACAATAATTCG	2756
Qy	320	---ThrIleLysGlyAlaLeu-----AlaMetProLeuMet 330	
			::: :::
Db	2757	GATACACTATAGTGAACGCTTTAAACGATGCTTTACCTATTATTC 2799	

```

RESULT 11
US-09-878-574-3629
: Sequence 3629, Application US/09878574
: Patent No. US20020110548A1
: GENERAL INFORMATION:
: APPLICANT: Byrum, Joseph R.
: APPLICANT: La Rosa, Thomas J.
: APPLICANT: Thompson, Michael D.
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
: TITLE OF INVENTION: plants
: FILE REFERENCE: 38-21(15401)B
: CURRENT APPLICATION NUMBER: US/09/878,574
: CURRENT FILING DATE: 2001-12-21
: PRIOR APPLICATION NUMBER: 09/333,535
: PRIOR FILING DATE: 1999-06-14
: NUMBER OF SEQ ID NOS: 15775
: SEQ ID NO 3629
: LENGTH: 411
: TYPE: DNA
: ORGANISM: Glycine max
: FEATURE:
: NAME/KEY: unsure
: LOCATION: (1)..(411)
: OTHER INFORMATION: unsure at all n locations
: OTHER INFORMATION: Clone ID: LIB3028-007-01-B1-H6
: US-09-878-574-3629

```

Alignment Scores:	
Pred. No.:	2,11e-08
Score:	.148,50
Percent Similarity:	49.15%
Best Local Similarity:	33.05%
Query Match:	8.11%
DB:	10
	2
Length:	4111
Matches:	39
Conservative:	19
Mismatches:	50
Indels:	11
Gaps:	2

US-09-857-613A-28 (1-350) x US-09-878-574-3629 (1-411)

Oy 128 TTPProlySerIleValaSpValaIGlyCYcSgLYlleGlyLySerAArgYrLeuAla 147
Db 3 TGGCCAAAAGTA- CTGAGATCTGTGGTGTGTACAGAGAGAGGTGCATTTCATGAGCT 60
Oy 148 LysLysPheGlyAlaThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAla 167
Db 61 GAAAAATTTTGATGTTGAGGTTGTGGCAATTCACCTCTCC-----ATA 102
Oy 168 Asn-----AlaLeuAlaAlaIacGlnLeuAlaAspLysValSerPheGln 183
Db 103 AACGTCATTTCTCTTGGCATTTGAACGTCCTATTGTGACCTCAATCTCTGGGTAGATTGGAA 162

Oy	184	ValAlaAspAlaLeuGlnGlnProPheSerTpsGlyGlnPheAspLeuValTlrPsrMet	203
Db	163	TetGGCGATTGCACATAAAAAACAATCCCTAGATACATTGTGTAACTATATCCCGT	222
Oy	204	GluserGlyGluHisMetProAspLysAlaLeuPheValGlyGluLeuAlaArgValAla	223
Db	223	GACGATTGTTCACATATAAAGATTAACCCACCACCTATTGAGATCATTTTACAAATGGTTG	282
Oy	224	AlaProGlyAlaIleIleIleIleValThrProCysHisArgAspLeuGlyPro	241
Db	283	AAGCGTGGAGGTACACTTATTATTACGATTATTCGCAAAAGTGAAGCAAGTCCA	336

```

RESULT 12
US-10-260-877-53
; Sequence 53, Application US/10260877
; Publication No. US20030021813A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Chovan, Linda E.
; APPLICANT: Hessler, Paul E.
; APPLICANT: Reich, Karl A.
; TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME
; TITLE OF INVENTION: SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: 'ESSENTIAL GENES'
; FILE REFERENCE: 6565.US.P1
; CURRENT APPLICATION NUMBER: US/10/260,877
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US/09/649,145
; PRIOR FILING DATE: 2000-08-25
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 53
; LENGTH: 753
; TYPE: DNA
; ORGANISM: H. influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(753)
; OTHER INFORMATION: HI-0095
US-10-260-877-53

```

Alignment Scores:	
Pred. No.:	4.02e-07
Score:	141.00
Percent Similarity:	44.68%
Best Local Similarity:	25.00%
Query Match:	7.70%
DB:	9
	Gaps: 5
	Length: 753
	Matches: 47
	Conservative: 37
	Mismatches: 74
	Indels: 30
	Gaps: 5

US-09-857-613A-28 (1-350) x US-10-260-877-53 (1-753)

```

QY 130 lysSerIleValAspValGlyCysGlyIleIleIleIleIleSerArgTyrIleuAlaIleValys 149
Db 118 AAAAAAGTGGAGAGCTTCCCTGTAAATGGGAGACAGCTGCAATGGATTGGCCAAACAA 177
QY 150 PheGlyAlaThrSerValGlyIleThrLeuSerProValGlnAlaIleArgAlaAsnAla 169
Db 178 TTTCCTGTGCATATTGAAGAGCTGTGATTGCATCAATCAAAATGCCGTACCAAAAGCAGCA 237
QY 170 LeuAlaIleAlaGlnGlyLeuAlaAspIleValSerPheGlnValAlaAspAlaLeuGln 189
Db 238 AATATTGAAGCAAAATGGCTTGCAGCAAAAAATTATGATGTACACGCTGGCAATGCCAATAG 297
QY 190 GlnProPheSerAspGlyGlnPheAspLeuValIleTrpSerMetGluSerGlyGlnHisMet 209
Db 298 TTGCCTTTCGAGATGAAGAAGTTTGTGATTTGTGCATCATGACAGCAAGCTCACAACTTTA 357
QY 210 Pro-----AspIleValIlePheValGlyGlnLeuAlaArgValAlaIleAlaProGlyAla 227
Db 358 CCCGCGAAGCGACAAAAAAAGCCATTTCACAAATATTTTTCGACGTGTTAAAAACCAATGCT 417
QY 228 IleIleIleIleValIleThrProCysHisArgAspLeuGlyProAspGluGlnIleSerLeuHis 247

```

Db 418 TTATGCTT-----ACTCAGCATGTATGCTGGCGGAATGATCATCAACTAT--- 468
QY 248 PRTPLGLUGlnAspLeuLeuLysLysIleCysAspAlaTyrTyrLeuProAlaTrpCys 267
Db 469 -----CTAGAAATAATATGCGCAAGCGATT----- 492
QY 268 SerThrSerAspTyrValLysLeuLeuGlnSerLeuSerLeuGlnAspIleLysSerGlu 287
Db 493 -----AAAGTGTACTGTACCGCATTAACGAAGAAT 522
QY 288 AspTrpSerArgPheValAlaProPheTrpProAlaValIleArgSerAlaPheTrp 307
Db 523 GGATGGAAAGCGATA-----TTCCAGAAAGTGTATTAGAAATGTGTACTTTC 573
QY 308 LysGlyLeuSerSerLeuLeuSer 315
Db 574 TCTGTGAGATGATCATTTACTTTC 597
RESULT 13
US-09-845-248-1
Sequence 1, Application US/09845248
Patent No. US20020042106A1
GENERAL INFORMATION:
APPLICANT: Nampoothiri, Madhavan
TITLE OF INVENTION: Nucleotide Sequences Which Code for the CMA Gene
FILE REFERENCE: 032301 WD 1161
CURRENT APPLICATION NUMBER: US/09/845,248
CURRENT FILING DATE: 2001-08-30
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1801
TYPE: DNA
FEATURE: Corynebacterium glutamicum
NAME/KEY: CDS
LOCATION: (248)..(1600)
OTHER INFORMATION:
US-09-845-248-1
Alignment Scores:
Pred. No.: 2,02e-06 Length: 1801
Score: 140.00 Matches: 83
Percent Similarity: 36.06 Conservative: 45
Best Local Similarity: 23.38 Mismatches: 120
Query Match: 7.658 Indels: 107
DB: 10 Gaps: 15
US-09-857-613a-28 (1-350) x US-09-845-248-1 (1-1801)
QY 53 LeuGlnGlnLysPro-----LysLysAspAsp 61
Db 674 ATGGAACAAACCCCTGCATGCGTAAAGCATGATCAACGGCGCTAGCATCCAGGACTCG 733
QY 62 LysLysLysLeuGlnLysGlyIleAlaGlnPheTyrAspGlnSerGlyLeuTrpGlu 81
Db 734 AAATCCCGCACAAGAACCATTAAGTACTACACTACAGCAGCGGCAATGACTTCTACTCC 793
QY 82 AsnIleTrpLysAspHisMetHis-----GlyPheTyrAspSerAspSerThrVal 99
Db 794 CTGTTTATGATGATTCATCCATCACTGCTGCGGCTATTATCAACAGCCAGATCA--- 850
QY 100 SerLeuSerAspHisArgAlaIleAlaGlnIleArgMetIleGlnGlnSerLeuArgPheAla 119
Db 851 AGTTTGGAAAGCCCAAGAAACAAATACCGCTCATCTTTGAAAGCTGCGCTG--- 907
QY 120 SerValSerGlnLysSerLysTrpProLysSerIleValAspValGlyCysGlyIle 139
Db 908 -----AAAGAGCGGATGCGCTCTCTAGACGTGGATGCGGTTGG 946
QY 140 GlyLysSerArgTyrLeuAlaLysPheGlyAlaThrSerValGlyIleThrLeu 159
Db 947 GGAGGATGTCGCTAC---GCCGCCAAACACGCTGTGAAGACCATGCGATTACGCTG 1003

QY 160 SerProValGlnAlaGlnArgAlaAsnAlaLeuAlaAlaGlnGlyLeuAlaAspLys 179
Db 1004 TCTGAACACCATATGAGTGGGCTACAGCAGATCAACAGCCAGTTTGAAGACCTTC 1063
QY 180 ValSerPheGlnValAlaAspAlaLeuGlnGlnProPheSerAspGlyLysPheAspLeu 199
Db 1064 GCGGAATTCCTTCATGATGATACCGCATGTTCGA-----GAACTGATTCGATGCG 1117
QY 200 ValTrpSerMetGlnSerGlyLysHisMet-----ProAsp----- 211
Db 1118 ATCTCAGCATATGGCATCTTGTAGACATCGGTGTAACAACACTATCCGACTTCTTGA 1177
QY 212 -----LysAlaLysPheValGlyLeuAla----- 220
Db 1178 TTGCTCAGCAGCAAACTCAAAACAGCGAGCATGATCTCAACACAGCATCACCCTACCA 1237
QY 221 -----ArgValAlaAlaProGlyAlaIleIleIleIleValThrTrpCysHisArgAsp 238
Db 1238 GACAAACCGCCCGCCAGCAGCATGCTATTAT---GATCGCTAC 1279
QY 239 LeuGlyProAsp----- 242
Db 1280 ATTTCCCGAGCGTGACTACTGCGCTCTGGCACCCTGATCAAGCATGACGAGACAC 1339
QY 243 -----GlnLysSerLeuHisPro 248
Db 1340 GTTTGGAGTGTGTCACGAGAAGAAACCTCCGCTTGTATTAACAGCACCCTGACGCG 1399
QY 249 TrpGlnGlnAspLeuLysLysIleCysAspAlaTyrTyrLeuProAla----- 265
Db 1400 TGCTGGAAACCTCAAGAAATTTGGAGAGAGCATGTAACCTCCCGGTGAACCCACT 1459
QY 266 -----TrpCysSerThrSerAspTyrVal 273
Db 1460 GCACGACTTTGGCTGATACGACAGTTCGGAATGGGATTTGCCCAACATCTGC 1519
QY 274 LysLeuGlnSerLeuSerLeuGlnAspIleLysSerGlnAspTrpSerArgPheVal 293
Db 1520 CACGTGACCAAGTACTGGTGTGA---CTGATGACGAGGAGCACTCGCGAGAA 1573
QY 294 AlaPro-----PheTrpProAlaValIleArg-SerAlaPheThrTrpLysGlyLeu 311
Db 1574 GTTCTGAAAGAAATGCTGACTATCTAAAGAACATGTTCTTTTAAAG----- 1625
QY 311 rSerLeuLeuSerSerGlyLysThrIleLysGlyAlaLeu 325
Db 1626 -----GTTCCCTAGGGGCGAGATTATTTAGTTAACTC 1661
RESULT 14
US-09-738-626-636
Sequence 636, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAMI, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988

; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: PatentIn ver. 3.0
 ; SEQ ID NO: 636
 ; LENGTH: 1353
 ; TYPE: DNA
 ; ORGANISM: Corynebacterium glutamicum
 US-09-738-626-636

Alignment Scores:

Pred. No.:	1.7e-06	Length:	1353
Score:	139.00	Matches:	63
Percent Similarity:	37.10%	Conservative:	42
Best Local Similarity:	22.26%	Mismatches:	122
Query Match:	7.60%	Indels:	56
DB:	9	Gaps:	12

US-09-857-613a-28 (1-350) x US-09-738-626-636 (1-1353)

```

QY 53 LeuGIUGlnLysPro-----LysLysAspasp 61
DB 427 ATGGAACAAAGCCCTGATGCGCTAAAGCAGCTGATCAGCGGCTAGCATCCAGGACTCG 486
QY 62 LysLysLysLeuGlnLysGlyIleAlaGluPheTyrAspLysSerGlyLeuTrpGlu 81
DB 487 AATCCCGGACAAAGAGGACCTAGCTACGACGATGCGGAGGAGTGGAGTCTACTCC 546
QY 82 AsnIleTyrGlyAspHisMetHis-----GlyPheTyrAspSerAspSerTrpVal 99
DB 547 CTGTTTGTAGATGATTCACAGACCTATACCTGCGCGATATTATCCACGCGCAATCA--- 603
QY 100 SerLeuSerAspHisArgAlaAlaGlnIleArgMetIleGlnLysLeuArgPheAla 119
DB 604 AGTTTGGAGAACCCCAAGAAACAAATACCGCTCATCTTTAAAAACGCTCTG--- 660
QY 120 SerValSerGlnLysSerTyrTrpProLysSerIleValAspValGlyGlyIle 139
DB 661 -----AAAGAGCGCATCCCTCTAGACGTCGATCGGTTGG 699
QY 140 GlyGlySerSerArgTyrLeuAlaLysLysPheGlyAlaIleHisSerValGlyIleTrpLeu 159
DB 700 GGAGGAGATGTCGCTAC---GCCGCCAAACAGGTGTGAACCCATCGAGTTACGCTG 756
QY 160 SerProValGlnIleAlaGlnArgAlaAsnAlaLeuAlaAlaIleGlnLysLeuAlaAspLys 179
DB 757 TCTGAACAGCAATATGAGTGGGCTCAAGCAGATCAAAACGCCAAGTTTGGAGACCTC 816
QY 180 ValSerPheGlnValAlaAspAlaLeuGlnInProPheSerAspLysGlnPheAspLeu 199
DB 817 GCGGAATTCGCTTCAATGATTCACCGGATGTTCA-----GAACTGATTCGATCGC 870
QY 200 ValTrpSerMetLysSerGlyGlnHisMet-----ProAspLysAlaLys 214
DB 871 ATCTGACGATCGCATATGTAACACATCGTGAACAACTATCCGACTCTTTGAA 930
QY 215 PheValGlyLeuLeuAlaArgValAlaAlaProGlyAlaIleIleIleIleValTrp 234
DB 931 TTCCTAGCAGCAACTCAACA-----GCGGAGCTAGTCTC----- 969
QY 235 CysHisArgAspLeuGlyProAspGlnLysSerLeuHisProTrpGlnLysLeu 254
DB 970 AACCAAGCATCACTACCCAGACACCGCCCGCCGAC-----GCA 1011
QY 255 LysLysIleCysAlaTyrTyrLeuPro-----AlaTrpCysSerThrSerAspTyr 272
DB 1012 GGTGCAATTAATGATCGTACATTTTCCCGAGGAGTGAACCTGAGCTGCGACCTG 1071
QY 273 ValLysLeuLeuGlnSerLeuSerLeuGlnAspLysSerGlnAsp 288
DB 1072 ATCAAGCAGATGACAGCAACGCTTTCGAGTCTGACGAGAAACCTCCGCTTGGAT 1131
QY 289 -----TrpSerArgPheValAlaIleProPheTrpProAlaValIle 301
  
```

DB 1132 TACCAAGCAGCCCTGACGCGTGTGCGAAACCTCAAGAAATTTGGAGAGACGATT 1191
 QY 302 ArgSerAla 304
 DB 1192 GAATCGCC 1200

RESULT 15

US-09-294-093B-5022

Sequence 5022. Application US/09294093B

Patent No. US20010051335A1

GENERAL INFORMATION:

APPLICANT: Laljudi, Raghunath, V.

APPLICANT: Ito, Laura, Y.

APPLICANT: Sherman, Bradley, K.

TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL

FILE REFERENCE: PL-0009 US

CURRENT APPLICATION NUMBER: US/09/294, 093B

PRIOR FILING DATE: 1999-04-16

PRIOR APPLICATION NUMBER: 60/082,567

PRIOR FILING DATE: April 21, 1998

NUMBER OF SEQ ID NOS: 6207

SOFTWARE: PERL Program

SEQ ID NO 5022

LENGTH: 290

TYPE: DNA

ORGANISM: Zea mays

FEATURE:

NAME/KEY: misc_feature

OTHER INFORMATION: Incyte ID No. US20010051335A1 700355594H1

US-09-294-093B-5022

Alignment Scores:
 Pred. No.: 3e-06
 Score: 128.00
 Percent Similarity: 35.42%
 Best Local Similarity: 33.73%
 Query Match: 6.99%
 DB: 10
 Gaps: 1

US-09-857-613a-28 (1-350) x US-09-294-093B-5022 (1-290)
 QY 130 LysSerIleValAspValGlyGlyIleGlyLysSerArgTyrLeuAlaLysLys 149
 DB 20 CAGAAGTACTGATGTCGATGCGATGGAATGAGAGGCGGACCTTTACATGCTGCAAC 79
 QY 150 PheGlyAlaHisSerValGlyIleTrpLeuSerProValGlnIleAlaGlnArgAlaAsnAla 169
 DB 80 TATGATGTCGATGCTTGTGATGATCTTCGCGTGAACATGCTT-----TCATTGCA 133
 QY 170 LeuAlaAlaIleGlnLysLeuAlaAspLysValSerPheGlnValAlaAspAlaLeuGln 189
 DB 134 ATGGAACGTGCTGATGAGCGCAAGTCTGTTGAATTCGAAGTTGCTGATGCGACCA 193
 QY 190 GlnProPheSerAspLysGlnPheAspLeuValTrpSerMetLysSerGlyGlnHisMet 209
 DB 194 AAGGATTACCAAAATATGTTTGAAGTCATCTACAGCGGTCGACACCATCTTCACATA 253

QY 210 ProAspLys 212
 DB 254 CAAGACAAG 262

Search completed: March 19, 2003, 09:06:49
 Job time: 100 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame-Plus-p2n model

Run on: March 19, 2003, 06:53:46 : Search time 1289 Seconds
(without alignments)

4397.533 Million cell updates/sec

Title: US-09-857-613a-28
Perfect score: 1830
Sequence: 1 MATVRIPTISCIHHTFRS.....IEGKKDLKFAITCKRPE 350

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09857613/runat.12032003.100036.6027/app_query.fasta.1.519
-DB=EST -QFMT=fastap -SUFFIX=trst -MINMATCH=0.1 -LOOPEXT=0 -LIST=45
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.ccl -LIST=45
-DOCALLIGN=200 -THR.SCORE=DCT -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=Pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09857613.ecgn.1.1.1456.ecgnat.12032003.100036.6027 -NCPU=6 -ICPU=3
-NO_XIPXY -NO_WMAP -LARGEOUTERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: EST:*
2: em_estha:*
3: em_esthum:*
4: em_estin:*
5: em_estinu:*
6: em_estlov:*
7: em_estpl:*
8: em_estro:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_iny:*
20: em_gss_pln:*
21: em_gss_vtc:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1136.5	62.1	1304	11 AY109799	AY109799 Zea mays
2	1029	56.2	599	14 BQ299297	BQ299297 sa045f11.
3	1024	56.0	669	10 BE331080	BE331080 so85a05.Y
4	980	53.6	567	14 BQ273543	BQ273543 sa024e11.
5	975	53.3	570	14 BQ786302	BQ786302 saq66c02.
6	971.5	53.1	778	12 BG127606	BG127606 EST473168
7	964	52.7	554	13 BM527813	BM527813 sal16bd10.
8	957	52.3	562	13 B1469324	B1469324 sal10d11.
9	946	51.7	556	13 B1469281	B1469281 sal109d01.
10	931	50.9	704	13 BQ288530	BQ288530 BJ288530
11	922	50.4	831	12 BG320098	BG320098 Zm03_01d1
12	899	49.1	664	13 B1463750	B1463750 BJ463750
13	884	48.3	515	14 BQ273302	BQ273302 sa023e12.
14	878	48.0	532	10 AM596072	AM596072 sl97e09.Y
15	869	47.5	628	9 A1487374	A1487374 EST245696
16	855	46.7	581	13 B1321569	B1321569 sal12h01.
17	844	46.1	509	12 BG511373	BG511373 sal17a10.
18	840	45.9	484	13 B1321583	B1321583 sal15a05.
19	836	45.7	708	14 BQ801716	BQ801716 WHE2817.G
20	824	45.0	503	12 BG790250	BG790250 sae67a12.
21	810	44.3	649	13 B1953869	B1953869 HVSME001
22	807	44.1	707	12 BQ416854	BQ416854 HVSME001
23	794	43.4	691	14 BQ116842	BQ116842 EST602418
24	788	43.1	540	10 BE331113	BE331113 so95906.Y
25	777	42.5	517	12 BQ046436	BQ046436 saas4a02.
26	775	42.3	587	10 BE340311	BE340311 EST344382
27	769	42.0	627	12 BG522228	BG522228 19-89 stre
28	763	41.7	708	12 BQ416864	BQ416864 HVSME001
29	759	41.5	439	12 BE804492	BE804492 sr80007.Y
30	754	41.2	444	13 BM094790	BM094790 sa121d02.
31	734.5	40.1	653	10 AW774108	AW774108 EST333338
32	728	39.8	553	12 BF631792	BF631792 NF008D03D
33	720	39.3	587	10 BE591261	BE591261 WHE1655-1
34	710	38.6	661	13 BJ447900	BJ447900 WHE1655-1
35	707	38.6	733	12 BF051091	BF051091 EST436266
36	685.5	37.5	533	9 A1489112	A1489112 EST247451
37	670	36.6	517	10 BE607043	BE607043 WHE0915.H
38	665	36.3	484	10 AM617931	AM617931 EST314005
39	662.5	36.2	730	12 BG523936	BG523936 37-16 ste
40	660	36.1	639	10 AV945401	AV945401 AY945401
41	658	36.0	491	9 A1485917	A1485917 EST244238
42	657	35.9	531	10 BE426891	BE426891 WHE0333.C
43	656	35.8	452	10 AM099184	AM099184 sd35h03.Y
44	650	35.5	623	12 BG447567	BG447567 NF004H01S
45	642	35.1	449	12 BG725508	BG725508 sae38c12.

ALIGNMENTS

RESULT 1
AY109799
LOCUS
DEFINITION
Zea mays CL364_1 mRNA sequence.
ACCESSION
AY109799
VERSION
AY109799.1 GI:21213657
KEYWORDS
SOURCE
ORANISM
Zea mays
Zea mays
HTE
Eukaryota: Viridiplantae: Streptophyta; Embryophyta: Tracheophyta;
Spermatophyta: Magnoliophyta: Liliopsida: Poales; Poaceae; PACC
Clade: Panicoideae; Andropogoneae; Zea.
I (bases 1 to 1304)
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.

REFERENCE

AUTHORS

TITLE Maize Mapping Project/Dupont Consensus Sequences for Design of Overgo Probes

JOURNAL Unpublished (2002)

REFERENCE 2 (bases 1 to 1304)

AUTHORS Coe, E.C.

TITLE Direct Submission

JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA

FEATURES Location/Qualifiers

Source 1..1304

/organism="Zea mays"

/db.xref="MaizeDB:631709"

/db.xref="taxon:4577"

/clone="Cl364.1"

/clone_id="Maize Mapping Project/Dupont Consensus Library"

/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed Dupont contigs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACS in conjunction with the Maize Mapping Project"

BASE COUNT 268 a 370 c 353 g 244 t 69 others

ORIGIN

Alignment Scores:

Pred. No.: 1.42e-106 Length: 1304

Score: 1136.50 Matches: 226

Percent Similarity: 71.87% Conservative: 32

Best Local Similarity: 62.95% Mismatches: 86

Query Match: 62.10% Indels: 15

DB: 11 gaps: 4

US-09-857-613a-28 (1-350) x AY109799 (1-1304)

QY 4 ValValaRgIlePrrrIleSerCysIleHis-----IleHisThrPheArgSer 20

DB 28 ATATCGCGGATCCCGCGCGCCCAATGCTACCGCGCGCTCCATTCG-----TCC 81

QY 21 GlnSerProArgThrPheAlaArgIleArgValGlyProArgSerTrpAlaProIleArg 40

DB 82 CACTCTCCAGAGACCTCCGAGCGCGCGCGCGCGAGCCACTACCGCGCTTCGCAC 141

QY 41 AlaseralAlaSerSerGluArgGlyIleValLeuGluGlnLysProLysLysAsp 60

DB 142 GTCCCGCGCCACTCCCGCGCTCCGACGCGCGCTCGTCCGATCGCGATCGCTCG 201

QY 61 Asp-----LysLysLysLeuLysGlyIleAlaGluPhe 72

DB 202 TCACAGCGCANN 261

QY 73 TyrAspGluSerSerGlyLeuTrpGluAsnIleTrpGlyAspHisMetHisIleGlyPhe 92

DB 262 TAGCAGAGATCGTCGGGGCTGTGGAGAACATCTGGGGCGCACCATGACACAGCGCTTC 321

QY 93 TyrAspSerAspSerThrValSerLeuSerAspHisArgAlaAlaGlnIleArgMetIle 112

DB 322 TAGGACTCCAGGAGAGCGCCCTCCATGACGCGCGCGCGCGCCAGATCGCATGATC 381

QY 113 GlnGluSerLeuArgPheAlaSerVal-----SerGluGluArgSerLysTrpProLys 130

DB 382 GAGGAGAGCGCTTCCTCCCGCGGTCCAGCGCTCAGATGANNNNNNNNNNNNNNNNNN 441

QY 131 SerIleValAspValGlyCysGlyIleGlyLysSerSerArgTrpLeuAlaLysPhe 150

DB 442 NNNNNAGTCGATGCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTC 501

QY 151 GlyAlaThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaAlaLeu 170

DB 502 GGAGCGCACTGCACGACGCGCTGACCGCTGTCAGACCGGAGAGAGAAATGCTCTC 561

QY 171 AAtaAlaGlnGlyLeuAlaAspLysValSerPheGlnValAlaAlaLeuGlnGln 190

DB 562 GCTGACGCGAGGGGTTGTGCGATCAGTTACTTCGAAAGTTCGATGCTCGAGCA 621

QY 191 ProPheSerAspGlyGlnPheAspLeuValTrpSerMetGluSerGlyGluHisMetPro 210

DB 622 CCGTTTCTGACGGGAGATGTCATGCTGTGTGCTCCATGAGAGATGGCGACATGCCG 681

QY 211 AspLysAlaLysPheValGlyGluLeuValArgValAlaAlaProGlyAlaIleIle 230

DB 682 GACAGAGAAATTTTGTGTAGCTAGCAGCGCGGGCGCTCCGAGGAGCAATATC 741

QY 231 IleValThrTrpCysHisArgAspLeuGluProAspGluGlnSerLeuHisProTrpGlu 250

DB 742 ATCGTACATGTCGTCATGAGAACTGATCCAGAACTGCTTAAGCCGATGAA 801

QY 251 GlnAspLeuLeuLysLysIleCysAspAlaTrpTrpLeuProAlaTrpCysSerThrSer 270

DB 802 CTGACCTTCCTAGAGAGATATGACCGCGCTACCTCCGAGCTGTGCTCACTTCA 861

QY 271 AspTrpValLysLeuGlnSerLeuSerLeuGlnAspIleLysSerGluAspTrpSer 290

DB 862 GACTTGTGACATTCGCAAGTCACTGTCTCGAGATATCAAGACGTCGACTGTCG 921

QY 291 ArgPheValAlaProPheTrpProAlaValIleArgSerAlaPheThrTrpLysGlyLeu 310

DB 922 GAGACGTGGCCCGCTTTGGCGCGCGCTGATTAATATGCGCTAACATGAGAGGCTTC 981

QY 311 SerSerLeuLeuSerSerGlyGlnLysThrIleLysGlyAlaLeuAlaMetProLeuMet 330

DB 982 ACCTCTGCTGACGACCGGATGAGACGATGAGAGCGCGGATGATGATGATGATGATG 1041

QY 331 IleGluGlyTrpLysAspLeuIleLysPheAlaIleIleIleIleIleIleIleIleIle 349

DB 1042 ATCCAGGCTCAAGAGAGGCGCTCATCAATCAATCAATCAATCAATCAATCAATCA 1098

RESULT 2

B0299297 599 bp mRNA linear EST 16-MAY-2002

LOCUS sa045f11.y1 Gm-cl054 Glycine max cDNA clone SOYBEAN CLONE ID: Gm-cl054-7197 5' similar to FR:09X1P9 09X1P9 GANMA-TOCOPHEROL METHYLTRANSFERASE. [1], mRNA sequence.

ACCESSION B0299297

VERSION B0299297.1 GI:20814819

KEYWORDS EST.

SOURCE soybean.

ORGANISM Glycine max

Ekaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 599)

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Rhanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Rheising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, T., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

Public Soybean EST Project

Unpublished (1999)

CONTACT: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available through: Resgen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact: cove@resgen.com web site: www.resgen.com

Seq primer: -40RP from Gibco

High quality sequence stop: 429.

Location/Qualifiers

1..599

```

/organism="Glycine max"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-c1054-7197"
/clone_lib="Gm-c1054"
/tissue_type="leaf, 3 week old, greenhouse grown"
/lab_host="DH10B"
/Note="Vector: Bluescript II SK-, Site_1: EcoRI, Site_2:
XhoI; The Harosoy NIL was constructed and seed was
provided by Dr. J. Specht, University of Nebraska
(Shoemaker and Specht, 1995). The cDNA library was
constructed from mRNA isolated from leaf tissue at various
developmental stages of 3 week old greenhouse grown
plants. Complementary DNA was synthesized from mRNA using
a primer consisting of a poly(dT) sequence with a XhoI
restriction site and a 3' anchor. EcoRI adapters were
ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into DH10B host cells (GibcoBRL). This
library was constructed in cooperation with Dr. Paul
Keim's laboratory at Northern Arizona University."
BASE COUNT      139 a      144 c      165 g      151 t
ORIGIN
Alignment Scores:
Pred. No.:      6,27e-96      Length:      599
Score:          1029.00      Matches:      199
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    56.23%      Indels:      0
DB:             14      Gaps:      0
US-09-857-613a-28 (1-350) x BQ299297 (1-599)
QY      6 ArgIleProThrIleSerCysIleHisIleHisThrPheArgSerGlnSerProArgThr 25
Db      3 AGGATCCCAACATCTCGATCGATCCACATCCACAGCTTCGTTCCATCCCTCGCAT 62
QY      26 PheAlaArgIleArgValGlyProArgSerTrpAlaProIleArgAlaSerAlaSer 45
Db      63 TTGCGCAACATCCGGGTGCGAGCCAGTCGTGGCTCCATTCGGGCAATCGGCAGAC 122
QY      46 SerGluArgGlyGluIleValLeuGlnGlnLysProLysAspAspLysLysLysLeu 65
Db      123 TCGGAGAGAGGGAGATGATGTTGAGCAGAACGCCAGAACAGATGACAAAGAACGCG 182
QY      66 GlnLysGlyIleAlaGluPheArgPheArgPheArgPheArgPheArgPheArgPheArg 85
Db      183 CAGAGAGGAGATCGCAGAGTTTACAGCAGCTTTCGCTTATGGAGAACATTGGGGC 242
QY      86 AspHisMetHisIleGlyPheArgPheArgPheArgPheArgPheArgPheArgPheArg 105
Db      243 GACGCAATGACCATGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 302
QY      106 AlaAlaGlnIleArgMetIleGlnGlnSerLeuArgPheAlaSerValSerGluGluArg 125
Db      303 GGTGTCAGATCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 362
QY      126 SerLysTrpProLysSerIleValAspValGlyCysGlyIleGlyLysSerArgTyr 145
Db      363 AGTAATATGCCCCAGAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 422
QY      146 LeuAlaLysLysPheGlyAlaThrSerValGlyIleThrLeuSerProValGlnAlaGln 165
Db      423 CTGGGCAAGAAATTTGGAGCAACACAGTGTAGCATCAGTCTGCTGCTGCTGCTGCTGCT 482
QY      166 ArgAlaAsnAlaLeuAlaAlaAlaGlnGlnLysLeuAlaAspLysValSerPheGlnAla 185
Db      483 AGAGCAAAATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 542
QY      186 AspAlaLeuGlnGlnProPheSerArgPheGlnGlnPheAspLeuValTrpSerMetGlu 204

```

```

Db      543 GACGCTACAGCAACATCTCTGACGCGCATTTGATCTGTGTGCTCATGAG 599
RESULT 3
BE331080
LOCUS
DEFINITION
BE331080 669 bp mRNA linear EST 04-DEC-2001
so95a05.y1 Gm-c1041 glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-c1041-1065 5' similar to TR:092SK1 092SK1 GAMMA-TOCOPHEROL
METHYLTRANSFERASE, mRNA sequence.
ACCESSION
BE331080
VERSION
BE331080.1 GI:9204856
KEYWORDS
SOURCE
ORGANISM
soybean.
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustoids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 669)
REFERENCE
Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Corryell,V., Khanna
,A., Bolla,B., Marr,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
,Y., Pearson,B., Swaller,T., Gibbons,M., Pape,D., Harrey,N., Schurk
,R., Riltter,E., Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
TITLE
JOURNAL
COMMENT
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact via email: ccutresgen.com
High quality sequence stop: 461.
FEATURES
source
1..669
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1041-1065"
/clone_lib="Gm-c1041"
/tissue_type="Senescing leaves, mature plant, greenhouse
grown"
/lab_host="DH10B"
/Note="Vector: pTV73Pac (Pharmacia); Site_1: EcoRI;
Site_2: HindIII; This library was constructed from mRNA
isolated from senescing leaf tissue of mature greenhouse
grown plants of the cultivar Williams. Complementary DNA
was synthesized from mRNA using a 3' anchored poly(dT)
primer. EcoRI adapters were ligated to the blunt-ended
cDNA fragments followed by digestion with EcoRI and
HindIII. The cDNA fragments were directionally cloned
into the EcoRI-HindIII restriction site of the pTV73-Pac
vector. The ligated cDNA fragments were transformed into
DH10B host cells. This library was constructed by Dr.
Randy Shoemaker."
BASE COUNT      154 a      156 c      184 g      174 t      1 others
ORIGIN
Alignment Scores:
Pred. No.:      2.37e-95      Length:      669
Score:          1024.00      Matches:      200
Percent Similarity: 97.10%      Conservative: 1
Best Local Similarity: 96.62%      Mismatches: 6
Query Match:    55.96%      Indels:      0
DB:             10      Gaps:      0
US-09-857-613a-28 (1-350) x BE331080 (1-669)
QY      1 MetAlaThrValAlaArgIleProThrIleSerCysIleHisIleHisThrPheArgSer 20

```

Db 49 ATGGCCACCGTGTGAGATCCCAACATCTCATGCATCCACATCCACACAGCTTCCTCC 108
 QY 21 GlnSerProArgThrPheAlaArgValGlyProArgSerThrPalaProIleArg 40
 Db 109 CATGCCCTCCGACCTTCCCGAGAAATCCGGGTCGGACCGAGGCTGGCTCTATTCCG 168
 QY 41 AlaSerAlaAlaSerSerGluArgGlyGluIleValLeuGluGlnIleProLysValAsp 60
 Db 169 GCATGGCAGCGAGCTCGAGAGAGGAGATATGATTGGACAGAGAGCGAAGAGCAT 228
 QY 61 AspLysLysLysLeuGlnLysGlyIleAlaGluPheTyrAspLysSerGlyLeuTrp 80
 Db 229 GACAAGAAACACTGCGAGAGGAAATCCGAGATTTCAGACAGCTTCCTGCTTATCG 288
 QY 81 GluAlaIleTrpGlyAspHisMetHisGlyPheTyrAspSerSerThrValSer 100
 Db 289 GAGAACATTTGGGGGACACACATGCACATGGCTTTATGACCTGGATTCACATGTTTCG 348
 QY 101 LeuSerAspHisArgAlaAlaGlnIleArgMetIleGlnIleSerLeuArgPheAlaSer 120
 Db 349 CTTTCGATCATCGTCTCTCAGATCCGATTCGATTCAGAGCTCTTGGCTTCTCT 408
 QY 121 ValSerGluGlnArgSerLysTrpProLysSerIleValAspValGlyGlyIleGly 140
 Db 409 GTTTCGTGAGACCGTAAATGCGCCAGAGATATGATGATGTTGGCTGCTGATANGT 468
 QY 141 GlySerSerArgTyrLeuAlaLysLysPheGlyAlaThrSerValGlyIleThrLeuSer 160
 Db 469 GGCACACTACGATCTGCGCCAGAAATTTGGACACACAGGTATGATCATCTGAGAGT 528
 QY 161 ProValGlnAlaGlnArgAlaAsnAlaLeuAlaAlaGlnIleGlyLeuAlaAspLysVal 180
 Db 529 CCGTTCAAGCTCATATAGACACATGCTTGTGCTGCTGCTCAAGGATTTGGCTATATGTT 588
 QY 181 SerPheGlnValAlaAlaAspAlaLeuGlnGlnIleProPheSerAspGlyGlnPheAspLeuVal 200
 Db 589 TCCCTTCATGTCGTGACCTCTACAGCAACCATTTCTGACGGCCAGCTTGTATGCTG 648
 QY 201 TrpSerMetGluSerGlyGly 207
 Db 649 TCGTTCATGACAGATGTCGAGAG 669

RESULT 4
 BQ273543 567 bp mRNA linear EST 07-MAY-2002
 LOCUS sa024e11.y1 Gm-cl054 Glycine max cDNA clone SOYBEAN CLONE ID:
 DEFINITION Gm-cl054-5206 5' similar to TR:09XTP9 09XTP9 GAMMA-TOCOPHEROL
 METHYLTRANSFERASE. [1] ; mRNA sequence.

ACCESSION BQ273543
 VERSION BQ273543.1 GI:20498613
 KEYWORDS EST.
 SOURCE soybean.
 ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 567)

TITLE Shoemaker R., Kelm, P., Vodkin, L., Eipelting, J., Corryell, V., Khanna
 'A', Bolla, B., Maria, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
 Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
 'Y', Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
 'R', Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McMan
 'R', Waterston, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 CONTACT: Shoemaker R./Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

This clone is available through: Resgen, Invitrogen Corp. 2130
 South Memorial Parkway Huntsville, AL 35801 For further information
 call: (800)-533-4363 or contact: ccuteresgen.com web site:
 www.resgen.com
 Seq primer: -40RP from Gibco
 High quality sequence stop: 420.
 Location/Qualifiers
 1. 567
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="SOYBEAN CLONE ID: Gm-cl054-5206"
 /clone.lib="Gm-cl054"
 /tissue.type="Leaf, 3 week old, greenhouse grown"
 /lab.host="DH10B"
 /note="vector: pBluescript II SK+, Site_1. EcoRI. Site_2:
 XhoI; The Harosoy NIL was constructed and seed was
 provided by Dr. J. Specht, University of Nebraska
 (Shoemaker and Specht, 1995). The cDNA library was
 constructed from mRNA isolated from leaf tissue at various
 developmental stages of 3 week old greenhouse grown
 plants. Complementary DNA was synthesized from mRNA using
 a primer consisting of a poly(dT) sequence with a XhoI
 restriction site and a 3' anchor. EcoRI adapters were
 ligated to the blunt-ended cDNA fragments followed by
 XhoI digestion. The cDNA fragments were directionally
 cloned into the EcoRI-XhoI restriction site of the
 pBluescript vector. The ligated cDNA fragments were
 transformed into DH10B host cells (GibcoBRL). This
 library was constructed in cooperation with Dr. Paul
 Kelm's laboratory at Northern Arizona University."

BASE COUNT 133 a 128 c 144 g 162 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 6,56e-91 Length: 567
 Score: 980.00 Matches: 187
 Percent Similarity: 99.47% Conservative: 0
 Best Local Similarity: 99.47% Mismatches: 1
 Query Match: 53.55% Indels: 0
 DB: 14 Gaps: 0

US-09-857-613a-28 (1-350) x BQ273543 (1-567)

QY 89 HISHISGLYPheTyrAspSerAspSerThrValSerLeuSerAspHisArgAlaGln 108
 Db 3 CACCATGGCTTTATGACCTCGGATTCACCTGTTCCCTTCGGATATGCTGCTCAG 62
 QY 109 IleArgMetIleGlnGluSerLeuArgPheAlaSerValSerGluGlnIleArgSerLysTrp 128
 Db 63 ATCCGATGATCCAGACAGCTCTTCCTTGGCTCTGCTTTCAGAGAGCGCTAATATGC 122
 QY 129 ProLysSerIleValAspValGlyGlyGlyIleGlyGlySerSerArgTyrLeuAlaLys 148
 Db 123 CCCAAGAGATATGTTGATGTTGGGTGCGCATATGAGGACGCTATGATACCTGCGCCAG 182
 QY 149 LysPheGlyAlaThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaGln 168
 Db 183 AATTTTGGACACACAGTATGACATCTGAGTCTGTTCAAGCTCAAGAGCAAT 242
 QY 169 AlaLeuAlaAlaGlnGlnIleLeuAlaAspLysValSerPheGlnIleValAlaAspAlaLeu 188
 Db 243 GCTCTTGGCTGCTGCTCAAGAGATTTGGCTATTAAGTTCTTTCAGAGTTGCTGACGCTTA 302
 QY 189 GlnGlnProPheSerAspGlyGlnPheAspLeuValTrpSerMetGluSerGlyGlnHis 208
 Db 303 CAGCAACATTTCTGACGCGCAGATTGATTCGCTGCTTCATGAGAGAGAGCAT 362
 QY 209 MetProAspLysAlaLysPheValGlyGlnLeuAlaArgValAlaAlaProGlyAlaIle 228
 Db 363 ATGCTGACAAAGCTTAGTTGTTGGAGATTAGCTGGGTGACACACAGGTGCTCACT 422
 QY 229 IleIleIleValThrTrpCysHisArgAspLeuGlyProAspGluGlnSerLeuHisPro 248

ORIGIN

Alignment Scores:	
Pred. No.:	2.84e-89
Score:	964.00
Percent Similarity:	99.46%
Best local Similarity:	99.46%
Query Match:	52.68%
DB:	13
Length:	554
Matches:	183
Conservative:	0
Mismatches:	1
Indels:	0
Gaps:	0

US-09-857-613A-28 (1-350) x BM527813 (1-554)

[illegible]

TITLE
JOURNAL
COMMENT

R., Waterston, R. and Wilson, R.
Public Soybean EST project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewaterston.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 for further information
call: (800)-533-4363 or contact via email: ccu@resgen.com
High quality sequence, stop: 421.

FEATURES

SOURCE

/organism="Glycine max"
 /db_xref="taxon:3847"
 /clone=GENOME SYSTEMS CLONE ID: Gm-cl053-2854"
 /clone_11b="Gm-cl053"
 /tissue_type="Whole seedling, 3 week old, greenhouse
 grown"
 /lab_host="DH10B"
 /note="vector: plusscript II SK+, Site_1: EcoRI, Site_2:
 XhoI; the Hairosoy NIL was constructed and seed was
 provided by Dr. J. Specht, University of Nebraska
 (Shoemaker and Specht, 1995). The cDNA library was
 constructed from mRNA isolated from whole seedlings of 3
 week old greenhouse grown plants. Complementary DNA was
 synthesized from mRNA using a primer consisting of a 5'
 poly(dT) sequence with a XhoI restriction site and a 3'
 anchor. EcoRI adaptors were ligated to the blunt-ended
 cDNA fragments followed by XhoI digestion. The cDNA
 fragments were directionally cloned into the EcoRI-XhoI
 restriction site of the plusscript vector. The ligated
 cDNA fragments were transformed into DH10B host cells
 (GibcoBRL). This library was constructed in cooperation
 with Dr. Paul Keim's laboratory at Northern Arizona

BASE COUNT
ORIGIN

130 a 136 c 153 g 143 t

Alignment Scores:	
Pred. No.:	1,52e-88
Score:	957.40
Percent Similarity:	99.47%
Best Local Similarity:	99.47%
Query Match:	52.30%
DB:	13
Length:	562
Matches:	186
Conservative:	0
Mismatches:	1
Indels:	0
Gaps:	0

US-09-857-613A-28 (1-350) x B1469324 (1-562)

[illegible]

```

|||||
Db 302 ATGATCCAAAGAGTCTCTCCCTTGTCTGAGGAGCGTAGTAATGCGCCAG 361
Qy 131 SerIleValAspValGlyCysGlyIleGlyGlySerSerArgTyrIleAlaLysLysPhe 150
Db 362 AGATATAGTATGATGGTGGTGGCATAGTGGCAGCTTCAATACCTGGCCCAAGAATTT 421
Qy 151 G1yAlaThrSerValGlyIleThrLeuSerProValGlnAlaGlnAlaAsnAlaLeu 170
Db 422 GGAGCAACCAAGTGTAGCATCAGCTGATCTTCAAGCTCAAGACCAATGCTCTT 481
Qy 171 A1aAlaIaGlnGlyLeuAlaAspLysValSerPheGlnValAlaAspAlaLeuGln 190
Db 482 GCTGCTGCTCAAGGATGGCGATAGATTCTTCAGGTTGCTGACGCTACAGCAA 541
Qy 191 ProPheSerAspGlyGlnPhe 197
Db 542 CCATTCTCTGACGGCAGTTT 562

RESULT 9
BI469281 556 bp mRNA linear EST 30-NOV-2001
LOCUS DEFINITION
BI469281.1 Gm-cl053 Glycine max cDNA clone GENOME SYSTEMS CLONE
ID: Gm-cl053-3025 5' similar to TR:09XIP9 09XIP9 GAMMA-TOCOPHEROL
METHYLTRANSFERASE. [1] ; mRNA sequence.
BI469281
BI469281.1 GI:15285390
EST.
soybean.
Glycine max
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;
Rosidae: eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 556)
Shoemaker, R., Kelm, P., Vodka, L., Erpelting, J., Coryell, V., Khanna
, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wyle, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers
, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
, R., Rittner, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntville, AL 35801 For further information
call: (800)-533-4363 or contact via email: ccl@resgen.com
High quality sequence stop: 422.
Location/Qualifiers
1. 556
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl053-3025"
/clone_id="Gm-cl053"
/tissue_type="Whole seedling, 3 week old, greenhouse
grown"
/lab_host="DH10B"
/note="Vector: pBluescript II SK+. Site_1: EcoRI; Site_2:
XhoI; The Harosoy NIL was constructed and seed was
provided by Dr. J. Specht, University of Nebraska
(Shoemaker and Specht, 1995). The cDNA library was
constructed from mRNA isolated from whole seedlings of 3
week old greenhouse grown plants. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with a XhoI restriction site and a 3'
anchor. EcoRI adapters were ligated to the blunt-ended
cDNA fragments followed by XhoI digestion. The cDNA

```

```

Fragments were directionally cloned into the EcoRI-XhoI
restriction site of the pBluescript vector. The ligated
cDNA fragments were transformed into DH10B host cells
(GibcoBRL). This library was constructed in cooperation
with Dr. Paul Keim's laboratory at Northern Arizona
University."
BASE COUNT 129 a 135 c 152 g 140 t
ORIGIN
Alignment Scores:
Pred. No.: 2,04e-87 Length: 556
Score: 946.00 Matches: 184
Percent Similarity: 99.46% Conservative: 0
Best Local Similarity: 99.46% Mismatches: 1
Query Match: 51.69% Indels: 0
DB: 13 Gaps: 0
US-09-857-613a-28 (1-350) x BI469281 (1-556)
Qy 11 SerCysIleHisIleHisThrPheArgSerGlnSerProArgThrPheAlaArgIleArg 30
Db 2 TCATGCATCCACATCCACACAGCTTCGCTTCCCAATCCCTCGACCTTGGCCGATCCGG 61
Qy 31 ValGlyProArgSerThrPalaProIleArgAlaSerAlaAlaSerSerGluArgGlyGlu 50
Db 62 GCGGACCCAGAGTCCGTGGCTCTTATTCGGCGCATCGGACGACGCTCGGAGAGGCGGAG 121
Qy 51 IleValLeuGlnGlnLysProLysAspLysLysLysLysLysLysLysLysLysLysLys 70
Db 122 ATGATATTGGACAGCAAGCCGAGAAAGATGACAGAAAGAACTCGAAGAGGAATCCGA 181
Qy 71 GluPheTyrAspGlnSerSerGlyLeuTyrGluAsnIleTyrGlyAspHisMetHisHis 90
Db 182 GAGTTTACGACAGACTCTTCTGCTTATGCGAGAACATTTGGGGGAGACATGACCAT 241
Qy 91 GlyPheTyrAspSerAspSerThrValSerLeuSerAspHisArgAlaAlaGlnIleArg 110
Db 242 GCTTTTATGACTCCGATTCACCTGCTTTCGATGATCATCGCTGCTCAGATCCGA 301
Qy 111 MetIleGlnLysSerLeuArgPheAlaSerValSerGlnLysArgSerLysTrpProLys 130
Db 302 ATGATCCAAAGAGTCTCTTCTGCTTCTGCTTCTGAGGACCGTAGTAATGCGCCAG 361
Qy 131 SerIleValAspValGlyCysGlyIleGlyGlySerSerArgTyrIleAlaLysLysPhe 150
Db 362 ACTATAGTGTGATGTTGGGTGGCATAGCTGCACGTCTAGATACCTGCGCAAGAATTT 421
Qy 151 G1yAlaThrSerValGlyIleThrLeuSerProValGlnAlaGlnAlaAsnAlaLeu 170
Db 422 GGAGCAACCAAGTGTAGCATCAGCTGATCTTCAAGCTCAAGACCAATGCTCTT 481
Qy 171 A1aAlaIaGlnGlyLeuAlaAspLysValSerPheGlnValAlaAspAlaLeuGln 190
Db 482 GCTGCTGCTCAAGGATGGCGATAGATTCTTCAGGTTGCTGACGCTACAGCAA 541
Qy 191 ProPheSerAspGly 195
Db 542 CCATTCTCTGACGGC 556

RESULT 10
BI288530 704 bp mRNA linear EST 09-APR-2002
LOCUS DEFINITION
BI288530 Y. Ogihara unpublished cDNA library, Wh-SL Triticum
aestivum cDNA clone whs114p22 5', mRNA sequence.
ACCESSION
BI288530
VERSION
BI288530.1 GI:20106970
KEYWORDS
EST.
SOURCE
bread wheat.
ORGANISM
Triticum aestivum
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae
; Triticaceae; Triticum.
REFERENCE
1 (bases 1 to 704)

```

AUTHORS	TITLE	JOURNAL	COMMENT
Ogihara, Y. and Murai, K.	Expressed genes in <i>Triticum aestivum</i> unpublished (2002)		Contact: Tadasu Shin-1

FEATURES

Source

BASE COUNT	169 a	192 c	194 g	148 t	1 others
------------	-------	-------	-------	-------	----------

ORIGIN

Alignment Scores:

Pred. No.:	9,74e-86	Length:	704
Score:	931.00	Matches:	173
Percent Similarity:	85.46%	Conservative:	21
Best local Similarity:	76.21%	Mismatches:	33
Query Match:	0.87%	Indels:	0
DB:	13	Gaps:	0

US-09-857-613A-28 (1-350) x BJ288530 (1-704)

Qy	83	ILETPPGLYASPHSMETHSHISGLPHERYTRSPSRASPSERTHVALSER	102
Db	22	ATCTGGGGCGAGCAATCCACACAGGGCTTCTACACATCCGGCGAGCGCCCTCCATATCC	81
Qy	103	ASPHISATGALAAAGLNILEARGMETILEGLINSERLEUAARGPHALASERVALSER	122
Db	82	GACCAACGGCCGGCCCAATCCGCATATATCAGAGAGGGCCCTCGGCTTCGGCGCGCTCCC	141
Qy	123	GLUGLIARGSERLYSTTRIPROLYSERILEVALASPYALGILCYSGILLEGILYLSER	142
Db	142	GACGATCCGACAACAAACCCAAAACGATTTTATGTCGGATCCGAAATCGCTGTATGC	201
Qy	143	SERAGTYRILEALALYSLYSPHGLYLATHRSERVALGILYILETHRLEUSERPROVAL	164
Db	202	TCMAATATCTCTGGCAACAAATATGTGACCAACATGCTCTGGATGCACATTCAGGCCAGTG	261
Qy	163	GLNALAGLNARGALASNALALEUALALALALAGNGLYLEUALAASPLYVALSERPHE	182
Db	262	CAGGCCGAGAGAGAAATGCGCTCTGCACACACAGGGGTTGTCTGGACAAGCCTCTTTC	322
Qy	183	GLNALALALASPALALEUGLNGLNPROHSESRASPSGLYNLPHASPLEUVALTRPSE	202
Db	322	CAAGTTGCGTAGCTCTGTGAGCAACATTTCTCTGTATGGCGCAGTTGATCTTGTCTGCTCT	381
Qy	203	METGLUSERGLYLHISMETPROASPSYALALYSPHEVALGLYLINLEUALAARGVAL	222
Db	382	ATGGAGAGTGGTGACACATATGCCCAACAAACAGAAATTGTTPAAGGACTCTGCACGGCTC	444

Qy	223	AlaIaIaProGlyAlaIleIleIleValaThrTrpCysHisArgAspLeuGlyProAsp	242
Db	442	GCAGCTCAGAGAGCAATCATATCATATCGTGAACCTGTGCCATAGGAACCTGGCCCAATCG	501
Qy	243	GIuGIuSerLeuHisProTrpGIuGIuAspLeuLeuLysLysIleCysAspAlaIaIyr	262
Db	502	GAGGACCTCACTGAAACCCTGACGAGCTGAATCTTTTGGAAAAAGATTGTGATGCATATTAAC	561
Qy	263	LeuProAlaIaTrpCysSerThrSerAspIyrAlaLysIleuLeuGIuSerLeuSerLeuGIu	282
Db	562	CTCCCGGATTGGTGTCTTCCTCGATTAATGTCAAGATATGCCGAACTCATTTGCTTTGAG	621
Qy	283	AspIleLysSerGIuAspTrpSerArgPheValaIaIaProPheTrpProAlaValaIleArg	302
Db	622	GATATCAAAACGGCGCATGTGCTGAAAAACGTGCCGCCGTTCTGTGCTGTATCCAA	681
Qy	303	SerAlaIaPheThrTrpLysGly	309
Db	682	TCAGCAGCTGACATGGAAAGCC	702

RESULT	11
BG320098/c	
LOCUS	BG320098
DEFINITION	Zm03_01d1_A Zm03-AAFC_ECCRC_cold_stressed_maize_seedlings Zea mays
ACCESSION	BG320098
VERSION	BG320098.1
KEYWORDS	EST.
SOURCE	Zea mays
ORGANISM	Zea mays

TITLE	Expressed Sequence Tags from Cold-Stressed Maize Seedlings Grown Under High Light Intensity
JOURNAL	Unpublished (2001)
COMMENT	Contact: Singh, J. A.

FEATURES

Source

```

/organism="Zea mays"
/cultivar="C0328"
/db_xref="taxon:4577"
/clone="zm03_01d11"
/clone_lib="zm03_AAFc_ECORC_cold_stressed_maize_seedlings"
/tissue_type="leaf, crown"
/dev_stage="4-leaf"
/note="Vector: Bluescript SR+/XhoI-ECOR1; Site.1: Eco RI; Site.2: Xho I; Corn seedlings at 4-leaf stage were exposed to low temperature/high light (100C/700-800uE/m2/s) for 4 days. Plants were grown/treated by J. Simmonds/L. Cass. library prepared by C. Piche using Stratagene kit."

```

BASE COUNT	166	a	237	c	213	g	214	t
ORIGIN								1 others

ORIGIN

Alignment Scores

Pred. No.:	1.02e-84	Length:	831
Score:	922.00	Matches:	174
Percent Similarity:	84.19%	Conservative:	23
Best Local Similarity:	74.36%	Mismatches:	37
Query Match:	50.38%	Indels:	0
DB:	12	Gaps:	0

US-09-857-613a-28 (1-350) x BG320098 (1-831)

```

Qy 116 LeuArgPheAlaSerValSerGluArgSerIleTrpProLysSerIleValAlaSpAl 135
Db 827 CTTCCCGCGTCCAGCCTCAGATATCCAGAGACACCAAAAAATAGTCATGCTTC 768
Qy 136 GlycSerGlyIleGlyGlySerSerArgTyrlLeuAlaLysLysPheGlyAlaThrSerVal 155
Db 767 GGATTCGATGCTGGTGGTACGCTCAAGTACTTGGCGAAGAAATACGAGCCAGTGCATC 708
Qy 156 GlyIleThrLeuSerProValGlnAlaGlnArgAlaAsnAlaLeuAlaAlaGlnGly 175
Db 707 GGGATCAGCTTGGACCCCTGTTCAAGCCGAGAGAGAAATGCTCTGCGTCGACGCGAGGG 648
Qy 176 LeuAlaAspLysValSerPheGlnValAlaAlaAspAlaLeuGlnGlnProPheSerSpLy 195
Db 647 TTGTCGGATCAGGTTACCTCCTCAAGTTCTGATGCTGAGCAACCGTTTCTTCAGCGG 588
Qy 196 GlnPheAspLeuValTrpSerMetGluSerGlyGlnHisMetProAspLysAlaLysPhe 215
Db 587 CAGTTCGATCTGCTGTGCTCCATGAGAGTGGCGAGCAGCATGCCGAGCAAGAGAAAGTTT 528
Qy 216 ValGlyGluLeuAlaArgValAlaAlaProGlyAlaAlaIleIleIleValThrTrpCys 235
Db 527 GTTACGAGCTAGACACGCGCTGCGCGCTCTGAGAGAGACAAATATCATGCTGACATGCTGC 468
Qy 236 HisArgAspLeuGlyProAspGluGlnSerLeuHisProTrpGluGlnAspLeuLys 255
Db 467 CATAGGAACCTGGATCCATCCGAAACCTGCTAAAGCCGATGAACTAGACCTCTCTGAGG 408
Qy 256 LysIleCysAspAlaTyrlTyrlLeuProAlaTrpCysSerThrSerAspTyrlValLysLeu 275
Db 407 AGGATATCGACAGCGCTACTACCTCCGAGCTGCTCACCCTCAGCATATGTGAACATT 348
Qy 276 LeuGlnSerLeuSerLeuGlnAspIleLysSerGluAspTrpSerArgPheValAlaPro 295
Db 347 GCCAAGTCACTGCTCTGCTGAGATATCAAGACAGCTGCTCGAGAACCTGCGACCGCCG 288
Qy 296 PheTrpProAlaValAlaArgSerAlaPheThrTrpLysGlyLeuSerSerLeuLeuSer 315
Db 287 TTTTGCCCGCGCTATTAATACAGCCCTAACATGMAAGGGCTTACCTCTGCTGACG 228
Qy 316 SerGlyGlnLysThrIleLysGlyAlaLeuAlaMetProLeuMetIleGluGlyTyrlLys 335
Db 227 ACCGATGGAAGACGATCAGAGCCGATGGATGGATCGCGCTAATGATCCAGGCTAACAG 168
Qy 336 LysAspLeuIleLysPheAlaAlaIleIleThrCysArgLysPro 349
Db 167 AAGGGCTCATCAATTCACCATCATCACTGTCGCAAGCCT 126

```

RESULT 12

LOCUS BJA63750 664 bp mRNA linear EST 23-MAY-2002
 DEFINITION BJA63750 K. Sato unpublished cDNA library, cv. Haruna Nijo
 germination shoots Hordeum vulgare subsp. vulgare cDNA clone
 bags30f10 5', mRNA sequence.

ACCESSION BJA63750
 VERSION BJA63750.1 GI:21142257
 KEYWORDS EST.

SOURCE Hordeum vulgare subsp. vulgare.
 ORGANISM Hordeum vulgare subsp. vulgare.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 ; Triticeae; Hordeum.

REFERENCE 1 (bases 1 to 664)
 Sato, K., Saitoh, D. and Takeda, K.
 TITLE Barley EST sequencing project in NIG and Okayama Univ
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856

FEATURES
 source
 1. 664
 /organism="Hordeum vulgare subsp. vulgare"
 /cultivar="Haruna Nijo"
 /db_xref="taxon:112509"
 /clone="bags30f10"
 /clone_lib="K. Sato unpublished cDNA library, cv. Haruna
 Nijo germination shoots"
 /tissue_type="shoots"
 /dev_stage="germination"

BASE COUNT 158 a 177 c 186 g 143 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.8e-82 Length: 664
 Score: 899.00 Matches: 168
 Percent Similarity: 86.24% Conservative: 20
 Best Local Similarity: 77.06% Mismatches: 30
 Query Match: 49.13% Indels: 0
 Db: 13 Gaps: 0

US-09-857-613a-28 (1-350) x BJA63750 (1-664)

```

Qy 73 TyrAspGluSerSerGlyLeuTrpGluAsnIleTrpGlyAspHisMetHisGlyPhe 92
Db 4 TACGACGAGTGTCTCCGGCTGTGGAGACATCTGGGCGAGACATCCACACCGCTTC 63
Qy 93 TyrAspSerAspSerThrValSerLeuSerAspHisArgAlaGlnIleArgMetIle 112
Db 64 TACGACCTCTGGGAGACCGCCCTCCATGCTCCGACACCCCGCGCCAGATCCGATGATC 123
Qy 113 GlnGluSerLeuArgPheAlaSerValSerGluGlnArgSerLysTrpProLysSerIle 132
Db 124 GAGGAGGCCCTGCTTCCGCGCCCTCCCGACCGCCGACCAACAAACCAATTT 183
Qy 133 ValAspValGlyCysGlyIleGlyGlySerSerArgTyrlLeuAlaLysLysPheGlyAla 152
Db 184 GTTATGTTGGATGGGAGATCGGCGGTAGCTCAAGTACCTGCGCAACAAATATGTGACA 243
Qy 153 ThrSerValGlyIleThrLeuSerProValGlnAlaGlnArgAlaAsnAlaLeuAla 172
Db 244 CAGTCTCTGGGATCACAATTGAGCCAGCTGCAAGCCGAGAGAGAAATCCCTACCGCG 303
Qy 173 AlaGlnGlyLeuAlaAspLysValSerPheGlnValAlaAspAlaLeuGlnProPhe 192
Db 304 GCACAGGGGTGGCAGACAGGCTTTTCCAGTTGCTGATGCTTGAGCAACCATTT 363
Qy 193 SerAspGlyGlnPheAspLeuValTrpSerMetGluSerGlyGlnHisMetProAspLys 212
Db 364 CCTGATGGGCACTTGTATCTTGTCTGTATGAGAGAGTGGAGACATGCCGACAAADA 423
Qy 213 AlaLysPheValGlyGluLeuAlaArgValAlaAlaProGlyAlaIleIleIleVal 232
Db 424 CAGAGTTTGTAGTGAGCTGCGACGCGCTCCAGCTCCAGGAGCAACATCATCTCTG 483
Qy 233 ThrTrpCysHisArgAspLeuGlnProAspGluGlnSerLeuHisProTrpGluLinsp 252
Db 484 ACCTGCTCCATAGGAGATGCTCCCTATCTGAGACTCTTGAACCTGAGACGCTGAT 543
Qy 253 LeuLeuLysLysIleCysAspAlaTyrlTyrlLeuProAlaTrpCysSerThrSerAsp 272
Db 544 CTTTGAAGAAAGATTGTGATCATATTAATCCCGGATGTGCTGCCCGCGGATAT 603
Qy 273 ValLysLeuLeuGlnSerLeuSerLeuGlnAspIleLysSerGluAspTrpSer 290
Db 604 GTCAAGATTGCGGAGCTCATTTGCTTGAGATATCAAAACGCGTGAAGTCT 657

```

RESULT 13

LOCUS BQ273302 515 bp mRNA linear EST 07-MAY-2002
 DEFINITION sa023e12.y1 Gm-cl054 glycine max cDNA clone SOYBEAN CLONE ID:

Gm-cl054-5207 5', similar to TR:09XIP9 09XIP9 GAMMA-TOCOPHEROL
METHYLTRANSFERASE. [1], mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORANISM

BQ273302
BQ273302.1 GI:20498372

EST.

soybean.

Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

Glycine.

1 (bases 1 to 515)

Shoemaker, R., Keim, P., Vodkin, L., Erpelidg, J., Corvett, V., Khanna,

A., Bolla, B., Matra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,

Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers,

Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk,

R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann,

R., Waterston, R., and Wilson, R.

Public Soybean EST Project

Unpublished (1999)

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available through: Resgen, Invitrogen Corp. 2130

South Memorial Parkway Huntville, AL 35801 For further information

call: (800)-533-4363 or contact: cou@resgen.com web site:

www.resgen.com

Seq primer: -40RP from Gibco

High quality sequence stop: 421.

Location/Qualifiers

1..515

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="SOYBEAN CLONE ID: Gm-cl054-5207"

/clone_11b="Gm-cl054"

/tissue_type="Leaf, 3 week old, greenhouse grown"

/lab_host="DH10B"

/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:

XhoI; The Harosoy NIL was constructed and seed was

provided by Dr. J. Specht, University of Nebraska

(Shoemaker and Specht, 1995). The cDNA library was

constructed from mRNA isolated from leaf tissue at various

developmental stages of 3 week old greenhouse grown

plants. Complementary DNA was synthesized from mRNA using

a primer consisting of a poly(dt) sequence with a XhoI

restriction site and a 3' anchor. EcoRI adapters were

ligated to the blunt-ended cDNA fragments followed by

XhoI digestion. The cDNA fragments were directionally

cloned into the EcoRI-XhoI restriction site of the

pBluescript vector. The ligated cDNA fragments were

transformed into DH10B host cells (GibcoBRL). This

library was constructed in cooperation with Dr. Paul

Keim's laboratory at Northern Arizona University."

Keim's laboratory at Northern Arizona University."

BASE COUNT 124 a 113 c 135 g 143 t

ORIGIN

Alignment Scores:

Pred. No.: 4.56e-81 Length: 515

Score: 884.00 Matches: 170

Percent Similarity: 99.42% Conservative: 0

Best Local Similarity: 99.42% Mismatches: 1

Query Match: 48.31% Indels: 0

DB: 14 Gaps: 0

US-09-857-613a-28 (1-350) x BQ273302 (1-515)

FEATURES

source

1..532

/organism="Glycine max"

/db_xref="taxon:3847"

LOCATION/Qualifiers

1..532

LOCUS

AM596072

532 bp

linear

EST 03-DEC-2001

DEFINITION

5197e09.y1 Gm-cl032 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:

Gm-cl032-353 5', similar to TR:09ZSK1 09ZSK1 GAMMA-TOCOPHEROL

METHYLTRANSFERASE. mRNA sequence.

ACCESSION

AM596072

GI:7283465

VERSION

AM596072.1

KEYWORDS

EST.

SOURCE

ORGANISM

soybean.

Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

Glycine.

1 (bases 1 to 532)

AUTHORS

Shoemaker, R., Keim, P., Vodkin, L., Erpelidg, J., Corvett, V., Khanna,

A., Bolla, B., Matra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,

Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers,

Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk,

R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann,

R., Waterston, R., and Wilson, R.

Public Soybean EST Project

TITLE

Unpublished (1999)

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available through: Resgen, Invitrogen Corp. 2130

South Memorial Parkway Huntville, AL 35801 For further information

call: (800)-533-4363 or contact: cou@resgen.com

Insert Length: 1413 Std Error: 0.00

High quality sequence stop: 429.

LOCATION/Qualifiers

1..532

```

/clone="GENOME SYSTEMS CLONE ID: Gm-c1032-353"
/clone.lib="Gm-c1032"
/tissue_type="Cotyledons of 8-day-old 'Williams' seedlings"
/lab_host="DH10B"
/notes="Vector: pBluescript II XR; Site 1: EcoRI; Site 2: XhoI; This cDNA library was constructed from mRNA isolated from cotyledons of 8-day-old 'Williams' seedlings which were propagated on paper towels with distilled water for 3 days (etiolated), then greenhouse grown for 5 days in potting soil. The cotyledons were flash-frozen in liquid nitrogen. StrataGene's cDNA Synthesis Kit (catalog number 200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA was hemimethylated. A modification of StrataGene's first-strand synthesis primer was used. An 'anchor' nucleotide (V=A, C, or G) was added to the 3' end of the primer [GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG(T)18V] to anchor the primer at the 5' end of the poly(A) tract. After second strand synthesis, the cDNA ends were filled in with cloned Pfu DNA, ligated to EcoRI adapters and subsequently phosphorylated. The cDNA was then precipitated and redissolved in sterile, RNase-, DNase-free water. The XhoI site within the first-strand synthesis primer was then restricted by digestion with XhoI from Promega (400U/ul); all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500 bp cutoff, using GIBCOBRL Life Technologies' cDNA size fractionation column. The column eluent was then precipitated, redissolved, and ligated into StrataGene's pBluescript II XR predigested vector (pBluescript II SK(+)) vector that has been digested with EcoRI and XhoI, and phosphorylated by StrataGene). 100% of the white and 87.5% of the blue colonies appear to contain recombinant plasmids with cDNA inserts, based on size (n=28 and 8 respectively). This library was constructed by Dr. Paul Kelm and Dr. Virginia Coryell."

```

BASE COUNT 128 a 131 c 146 g 124 t 3 others

ORIGIN

Alignment Scores:

Pred. No.:	1.97e-80	Length:	532
Score:	878.00	Matches:	171
Percent Similarity:	98.28%	Conservative:	0
Best Local Similarity:	98.28%	Mismatches:	3
Query Match:	47.98%	Indels:	0
DB:	10	Gaps:	0

US-09-857-613A-28 (1-350) x AWS96072 (1-532)

QY 1 MetalaThValaValaRgIleProThrIleSerCysIleHisIleHisThrPheArgSer 20
 |||||
 DB 11 ATGGCCACCGTGTGAGATCCCAACATCTCATCATCATCATCATCATCATCATCATCAT 70
 |||||
 QY 21 GlnSerProArgThrPheHlaArgIleArgValGlyProArgSerTrpAlaProIleArg 40
 |||||
 DB 71 CAATCCCTCGCATCTTCCCGCAAGATCCGGGTGGACCCAGGTGGGCTCTCTATTTCG 130
 |||||
 QY 41 AlaseralAlaIleSerSerGluArgGlyIleValIleGluIleGlnIleProIleValSer 60
 |||||
 DB 131 GCATCGGCGAGCGAGCTCGAGAGAGGAGATAGTATTGGACAGAGAGCGAAGAGAT 190
 |||||
 QY 61 AsplysLysLysLeuGlnLysGlyIleAlaGluPheTrpAspLysSerGlyLeuTrp 80
 |||||
 DB 191 GACAG 250
 |||||
 QY 81 GluAsnIleTrpGlyAspHisMetHisGlyPheTrpAspSerSerPheValSer 100
 |||||
 DB 251 GAGAACATTGGGCGACCATGACCCATGCTTTATGACCTCGGATTCCTGCTTTCG 310
 |||||
 QY 101 LeuSerAspHisArgAlaIleGlnIleArgMetIleGlnIleSerLeuArgPheAlaSer 120
 |||||

```

DB 311 CTTTCGATCATCTGCTCTCTCATCATCCAGATCATCTCTTCTTCTTCTCTCT 370
QY 121 ValSerGluGluArgSerLysTrpProLysSerIleValAlaValGlyCysGlyIleGly 140
|||
DB 371 GTTTCGAGAGAGCGTAGTAATGCGCCAGAGAGATAGTATTGATTTGGGTGGCATANCT 430
QY 141 GlySerSerArgTrpLeuAlaLysLysPheGlyAlaThrSerValGlyIleThrLeuSer 160
|||
DB 431 GCGAGCTTAGACTCTGACCAGAAATGTAAGAGCAACAGTGTAGGCACTCTGACT 490
QY 161 ProValGlnAlaGlnArgAlaAsnAlaLeuAlaAlaAlaGln 174
|||
DB 491 CCGTTTCAGCTCAAGACCAAGCAATGCTCTTCTGCTGCTCTCA 532

```

RESULT 15

AI487374 628 bp mRNA linear EST 18-MAY-2001

LOCUS

DEFINITION EST245696 tomato ovary, TAMU Lycopersicon esculentum cDNA clone

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Lycopersicon esculentum

tomato.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; eusterids I; Solanales; Solanaceae; Solanum;

Lycopersicon.

1 (bases 1 to 628)

Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Alang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Niemman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J.

Generation of ESTs from tomato carpel tissue

Unpublished (1999)

CONTACT: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>.

FEATURES

source

1..628

/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db_xref="taxon:4081"

/clone="CLEDI3C24"

/clone.lib="tomato ovary, TAMU"

/tissue_type="carpel"

/dev_stage="5 days pre-anthesis to 5 days post-anthesis"

/lab_host="XLI-Blue MRF"

/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; cLED - Tomato Carpel EST library. OligodT-primed and directionally cloned cDNA in vector Lambda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."

BASE COUNT 170 a 133 c 160 g 165 t

ORIGIN

Alignment Scores:

Pred. No.:	2.07e-79	Length:	628
Score:	869.00	Matches:	164
Percent Similarity:	88.83%	Conservative:	19
Best Local Similarity:	79.61%	Mismatches:	23
Query Match:	47.49%	Indels:	0
DB:	9	Gaps:	0

US-09-857-613A-28 (1-350) x AI487374 (1-628)

QY 103 AspHisArgAlaIleGlnIleArgMetIleGlnIleSerLeuArgPheAlaSerValSer 122
 |||||
 DB 9 GATCATCTGCTGCTCAGATCGTAGATTGAACAGGCTCAAGTTTCTCTCTATTCT 68
 |||||
 QY 123 GluGluArgSerLysTrpProLysSerIleValAlaValGlyCysGlyIleGlySer 142
 |||||

